

学校的理想装备

电子图书·学校专集

校园网上的最佳资源

中小学信息科学知识

信息存贮与检查



## 第一章 信息引论

### 第一节 信息的涵义

信息作为名词术语，在中国，最早出现在唐代诗人李中的诗句中：“梦断美人沉信息，目穿长路倚楼台”。但是，信息作为一个科学概念以及科学研究对象，却只有四五十年的历史。然而，在人类社会的演变和发展过程中，信息一直在积极地发挥着人类已经意识或还没有意识到的重要作用。

圣经中记载，在上古时期，洪水泛滥，诺亚自造方舟，避免了灭顶之灾。方舟在洪水中飘荡许久，诺亚想知道洪水是否已退，就放出飞鸽，待飞鸽衔回一支橄榄枝，诺亚根据飞鸽的返回时间及新嫩的绿枝这样的信息，推断洪水已退。和平来临了，人类获救了。这是人类利用信息的最早的文字记载。

自古以来，人们随时都在自觉不自觉地接收、传递、存贮和利用信息。人类的信息活动也从来没有间断过。

20世纪科学技术的空前进步，使人们对信息的认识和研究逐步建立在科学的基础上，形成了以信息为研究对象的理论——信息论，并迅速发展为一门具有丰富内涵的新学科——信息科学。

1948年，美国科学家申农（C.Shannon）和维纳（N.Weiner）先后发表了《通信的数学理论》和《控制论：或动物和机器中控制和通信的科学》两篇著作，为信息学的建立奠定了理论基础。

然而，对于信息的定义，人们还没有一致的认识。

申农这样描述信息：“信息是用以消除随机不确定性的东西”，从信息具有减少人们认识的不确定性的功能上概括了信息的特征。

维纳认为：“信息就是信息，不是物质，也不是能量”，把信息看成是与物质、能量具有同等重要性的客观世界三大要素之一。

我国著名的科学家钱学森称信息是“激活了、活化了的知识”，认为信息“就是为了解决一个特定的问题所需要的知识”，对信息进行了科学的概括。

可见，信息具有极其丰富的涵义，概括起来，信息的概念有三点是共同强调的：

1. 为了一定目的而系统收集起来的正确的事实所构成的知识，也就是说，经过传递而又为人们所吸收的知识构成信息。

2. 通过传递而进入人类社会交流体系的运动着的知识构成信息，即一方面信息可以通过一定的物质载体或其它交流形式进行传递，另一方面获取信息必须经过传递。

3. 信息不仅是传递中的知识，而且必须是有效的知识。凡是人类社会的信息均构成一定的效用，或与接收者原来知识状态吻合而被认可；或有相异，从而发生知识的重构，生成新的知识结构。

因此，可以说，在社会交流过程中，知识变为信息，信息转化为知识，它们总是处在相互作用之中。知识是静态的，以实践为基础，通过抽象思维，对客观事物规律性给予概括。知识信息是人类社会中客观存在着的，以知识形态为主，同时包括数据、新闻、消息等非知识形态的社会信息。知识信息是知识的激活，知识是知识信息的条理化。

其中有一点是可以肯定的，即通常听说的原始数据并不能称之为信息。

原始数据和信息的主要区别就在于：一切信息虽然都是由数据组成的，但并非一切数据都能产生消除不确定性而导致有用的信息。

原始数据需要经过加工处理而转变为信息，经过交流和传递实现信息应用。

因此，只有科学地了解和认识信息，才能更好地把握信息，才能使信息更好地为科学技术、经济和社会发展服务。

## 第二节 信息的特性和作用

### 一、信息的特性

信息之所以区别于物质与能量，并具有与物质、能量同等的重要作用，是源于信息所拥有的特性。物质在使用中是消耗的；能量就其个体而言在使用中也是消耗的，就其整体而言则是永恒的；而信息在其传递和使用过程中，可以重复使用，并可通过信息的加工处理而产生信息增值。信息作为一种资源，主要具有以下特性：

1. 客观性。信息的存在是客观的，它导源于客观存在的物质及其运动两大特性。从有人类存在以前，乃至今天，人类以外的各种生物就利用着大自然无穷无尽的信息资源。信息的客观性还表现为它是以物质的客观存在为前提的，即使是主观信息，如决策、判断、指令、计划等，也有它的客观实际背景，并以客观信息为“原料”，受客观实践的检验。

2. 依附性。信息总是依附于一定的物质载体而存在，需要某种物质承担者。如果不依附各种适当的载体，信息的含义和价值则不能传递和发挥。声音、语言、文字、颜色、图像、各种符号、光电磁、生物等各种信息信号、纸张、胶片、磁带光盘、人的大脑等等，无一不是信息的载体。

3. 可传递性。信息的产生就是同信息的传递联系在一起的，信息在传递过程中发挥它的作用。信息的传递和流通过程是一个重复使用的流通过程，在这一过程中，信息的占有者不会因传递信息而失掉信息，一般说来，也不会因多次使用而改变信息的自身价值。

信息在时间上的传递通常被称之为信息的存贮。

4. 可塑性。信息可以加工处理，可以压缩、扩充和叠加，也可以变换形态。在流通和使用过程中，经过综合、分析、再加工，原始信息可以变成二次信息和三次信息；原有的信息价值也可以实现增值；为了有效地交流和传递，借助于先进的信息技术，文本、图像、数字、语言等各种形态的信息均可实现互相转换。

5. 时效性。现代社会中，信息的使用周期迅速缩短，信息的价值实现取决于及时地把握和运用信息。信息是活跃的，不断变化的，及时地获取有效的信息将获得信息的最佳价值，如时效性很强的天气预报、经济信息、交易信息、科学信息等。不能及时地使用最新信息，信息的价值就会随其滞后使用的时差而减值或贬值。

6. 共享性。信息的共享性主要表现在同一内容的信息可以在同一时间由两个或两个以上的使用者使用，而信息的提供者并不失去所提供的信息内容和信息量，它是信息资源的广泛提供与利用的基础。

### 二、信息的基本作用

作为一种知识交流和社会交流，信息在人类社会和科技发展具有重要

作用。

1. 信息是人类认识客观世界及其发展规律的基础。信息的基本功能，主要表现为信息的认识功能。它是辩证唯物主义认识论的基础，是揭示客观事物发展规律的重要途径。信息是客观事物及其运动状态的反映，客观世界里到处充满着各种形式和内容的信息，人类的认识器官，包括感觉器官和思维器官，对各种渠道的信息进行接收，并通过思维器官将已收集到的大量信息进行鉴别、筛选、归纳、提炼、存贮而形成不同层次的感性认识和理性认识。在这一认识过程中，人类是认识论的主体，信息是认识论的客体。

2. 信息是科学研究的必要条件。人类的知识具有继承性和共享性。科学的大厦是千万个科学家在历史的进程中逐步建立起来的，任何一位科学家在从事科学研究时，都不能不借鉴前人的成果和依靠同时代其他人的帮助，这就是说，他同时需要在时间上和空间上的信息传递。在这个问题上，自然科学家和社会科学家各具特色。自然科学研究是一个比较严格的循序渐进过程，其每一个成果都是在前人建立的基础上发展起来的，由于自然科学研究的抽象性、准确性和严格性，独立研究是非常困难的，因此，信息交流就十分必要，并且对信息的数量和质量等都有较高的要求。对于社会科学而言，其研究成果不具有像自然科学研究成果那样的准确性，但在多样性上远远超过前者，并更多地受社会政治和人为因素的影响，其不同观点和理论的统一，较之自然科学也往往更困难和更需要时间，因此，需要更多地占有信息资源，充分地认识和把握各种不同的信息观点。

概括地说，信息在科学研究工作中的具体作用如下：

第一，掌握动态，选准课题，避免科研工作的重复浪费；

第二，掌握“已知”，利用他人成果，加快科学研究进程；

第三，提高科学劳动生产率，节约科研经费和工作投资。

3. 信息是管理和决策的主要参考依据。从广义上讲，任何管理系统都是一个信息输入、变换、输出的信息与信息反馈系统。这是因为，管理者首先要知道被管理对象的一些基本情况。在一定程度上消除对被管理对象认识的不确定性后，制定相应的对策，进而实施管理。更进一步讲，任何组织系统要实现有效的管理，都必须及时获得足够的信息，传输足够的信息，产生足够的信息，反馈足够的信息。只有以一定的信息为基础，管理才能驱动其运行机制，只有足够的信息，才能保证管理功能的发挥。

从某种意义上讲，信息都是为决策服务的，是为人们未来行动服务的。正确的决策与多种因素有关，如决策体制、决策方法、领导者的能力和素质等，但决定性的因素，还是取决于对客观实际的了解，对未来行动及其后果的正确判断，而正确的判断又主要依赖于全面、及时和准确的信息。

4. 为国民经济的建设和发展服务。信息作为一种资源，可以创造财富，通过直接或间接参与生产经营活动，为国家经济建设的各个方面发挥出重要的作用。

作为一种知识，信息的价值也是无法直接计算的，但它的经济效益却是实实在在的。一项适时对路的信息，可以带来一种新产品，或在贸易谈判中处于有利地位。信息的交流可以鼓励竞争，消除垄断，使不同的企业或工程项目得到相互促进的发展。技术经济信息可以有利于产品的更新换代，质量的提高，促进技术的进步和生产的发展；市场信息能提高全民经济生产的协调性；等等。在我国当前经济体制改革的新时期里，信息工作应把为国民经

济建设服务作为自己的重要方针之一。

### 第三节 信息源

信息源是指信息传递过程中的信息发送端或生成端的总称，它是信息产生的源头。

广泛存在着的多种类型的信息源，既是信息工作的基础，又是信息工作的对象。对信息源的类型、形态、结构及特点的分析和研究，是信息存贮、检索和开发利用的重要课题。

#### 一、信息源的类型

信息源的类型可依不同的标准划分如下：

1.按信息的保存性，可分为正式记录的信息源、非正式记录的信息源。前者指的是以可以保存的形式记录的信息源，如各种印刷品、缩微、声像、机读载体的文档、资料或出版物；后者指的是没有正式记录、无法保存的信息源，如会议、电话、口头交流等。

2.按信息源产生的时间顺序可分为先导信息源、实时信息源、滞后信息源。先导信息源是指产生时间先于社会活动的信息源，如天气预报、市场预测等；实时信息源是指在社会活动过程中产生的信息源，如实验记录、产品、讲座或报告等；滞后信息源是指某一社会活动完成之后产生的反映这一活动的信息源，如报刊、会议论文等。

3.按信息源的生产过程可分为原始信息源、加工信息源。原始信息源即为一次信息源，它是人类社会实践活动中直接产生或得到的各种数据、概念、知识、经验及其总结。加工信息源则是有关单位根据社会的不同需求对原始信息源进行加工、分析、改编、重组，生产人们社会活动所需的各种信息源。加工信息源又可按其加工的方式和深度的不同分为二次信息源、三次信息源。

4.按信息源产生信息的内容又可分为社会科学信息源、自然科学信息源和科技信息源。

#### 二、文献信息源

凡是人类的知识用文字、图形、代码、符号、声频、视频等方式和技术手段记载在一定载体上的每一件记录，统称为文献。简言之，文献是记录着知识或信息的物质载体。其中，知识、信息是文献的实质内容、灵魂；物质载体是知识、信息存贮、传递的主要工具和外在形式；文字、图形、符号等都是记录或表达知识、信息的手段，是无形的知识、信息与有形的物质载体的联系物。

习惯上，我们把记录科学知识的每一份物质载体称为科学文献，而把科学文献的汇总称为科学文献流。

在各类信息源中，文献是最主要、最常用的基本信息源。

##### （一）文献信息的主要形式

文献信息根据其载体的物质形态，基本上可分为手抄型、印刷型、缩微型、机读型和视听型。

1.手抄型。主要是指古旧文献和未经付印的手稿及技术档案之类的资料，其中可供开发利用者颇多。

2.印刷型。属传统的印刷形式，以纸张为载体，通过包括铅印、石印、

油印、胶印等方式得来的产品。此类文献历史悠久，收存丰富、系统，因其方式灵活、方便、广泛，保存时间相对较长，使用便捷。但其缺点是出版速度慢，体积大、笨重，收藏空间大，保管不易，且纸质易老化碎变。

3. 缩微型。即缩微复制品，它以感光材料为载体，利用摄影技术将手抄型或印刷型文献缩摄而形成的文献形式，包括缩微胶卷、缩微平片、缩微胶套和幻灯片等。缩微文献体积小、存贮密度高、存贮容量大、价格便宜，便于保存和检索。但其缺点是必须借助于阅读机才能阅读。

4. 机读型。采用一种利用机器——电子计算机才能进行阅读的新型载体，即采用电子计算机和光电磁技术，通过编码和程序设计，把文字信息转换成计算机可读的语言，输入机器，存贮于磁带、磁盘等载体。阅读时，由计算机按指令和存入的标识将存入的信息转换成文字或图像输出。机读型文献因存贮容量大，传输速度极高，可存贮各种形式的信息，宜用来做大量的情报信息存贮和快速的文献信息检索。

5. 视听型。又称声像资料，它通常以感光材料或磁性材料为载体，以光学感光或电磁转换为记录手段而产生出来的一种文献。它包括录像带、录音带、幻灯片、唱片、科技电影等，由于它运用录音、录像和摄影等技术直接记录声音和图像，所以提供的图像、声音逼真，宜于记载难以用文字表达和描绘的形象资料和声频资料，如用以记载野生动物保护区珍稀禽兽的活动及其吼鸣声。通过播放，可达如见其形，如闻其声的真切效果，给人以直感的感觉。

## （二）文献信息的结构层次

信息工作的主要任务是对信息进行深层次开发和综合利用，为了有效地发掘出文献的信息内容，必须对文献进行一定的加工，因此，根据文献的产生次序和加工整理的程度不同，可将文献划分为四个层次结构：

1. 零次文献。也称零次信息。指未经正式发表或不宜公开和大范围内交流的比较原始的素材、底稿、手稿、书信、工作文稿、工程图纸、考察记录、实验记录、调查稿、原始统计数字，以及各种口头交流的知识、经验或意见论点等，此类更为原始性的文献多保留于科技人员之手。另外，科技部门、有关管理部门和计划部门也有收藏。这类文献在较小的范围内交流、使用、参考，其传播渠道少，或常保密，或限制使用以及因珍稀的原因，不为人们知晓而多被埋没。其形式多是抄件、打字件、油印件、内部铅件、复制印刷、描图以及内部录音、录像等等。其特点是信息来源直接、真实，内容新颖。

2. 一次文献。习惯上称作原始文献，也称原始信息源，它是作者依据本人的科研和工作成果而形成的文献，这类文献是脑力劳动的正式产品，是科研成果的一种主要表述方式，代表新知识，组成了可供交流的系统性信息。此类文献主要包括：图书、期刊和报纸、科学考察报告、研究报告、会议论文、学位论文、专利说明书、技术标准、政府出版物、产品样本等。

一次文献有如下特点：

（1）创造性：一次文献是人们根据自己在生产和科学研究中的成果撰写的，是创造性劳动的结晶。它报道新成果、新技术、新发明、新创造。例如专利，它具有新颖性和创造性，反映了发明创造、技术革新与改进的创造性劳动成果。正由于一次文献的创造性，故受到人们重视。

（2）原始性：一次文献是一种原始的创作，也必然是初次发表的，一般是作者根据自己所积累的原始素材、原始数据创作而成，故其既有真实可靠

的一面，又有特定性和不成熟一面。

(3) 多样性：一次性文献是每个作者的不同成果，故在内容上多样化；另外，在表现形式上也呈现出多样性，有文学读物、期刊论文，也有研究报告、学位论文等等。

3. 二次文献。指根据实际需要，按照一定的科学方法，将特定范围内的分散的一次文献进行加工整理使之有序化而形成的文献。它能较为全面地系统地反映某学科某专业文献的线索，是检索一次文献的工具。这类文献有：期刊（仅限于揭示和评述一次文献的那些类型）、索引和文摘、论文集、图书馆目录以及参考书中的百科全书、专科全书、手册、大全、字典、词典、表格、图谱、数据等。

二次文献具有以下特点：

(1) 集中性：二次文献集中了某个特定领域范围的文献。它可以是某个信息部门的所有书刊资料，某个学科领域的文献，某个作者的所有文献等。二次文献是在所集中的某个特定范围的文献基础上，用科学的方法加工整理、组织编排而成，它比较完整地反映了某信息部门、某学科、某作者等的文献情况。

(2) 工具性：二次文献可称之为工具性的文献，它以特定的方法，简练的语言揭示文献的外部特征和内容特征，并加以科学的编排。它是累积、报告和查找文献线索的一种工具。一般所说的信息检索，即指对于此类文献的有效利用，从中查检到一定的知识信息或某项课题的文献线索。

(3) 系统性：二次文献本身具有自己的系统结构，为了方便利用，一般提供多个检索途径。所以一种好的二次文献往往由几个部分组成，具有比较固定的体系结构。另外，二次文献在总体上也有自己的系统性，特别是检索刊物体系，因为各学科的文獻都有自己的检索工具，做到不重复且没有重大遗漏，就必须统筹规划，分工协作，建立科学的检索刊物体系，且各个检索刊物也必须保持自己的连续性和系统性。

4. 三次文献。指通过二次文献提供的线索，选用一次文献的内容，进行分析、综合、研究后而编成的文献。一般包括专题述评、动态综述、进展报告、学科年度总结等。

三次文献具有以下特点：

(1) 综合性：三次文献是在大量有关文献的基础上，经过综合、分析而成。综合性就是将大量分散的有关特定课题的文献、事实和数据进行综合、评价、筛选，以简练的文字扼要叙述出来，内容十分概括。它可以是纵向综合，如某学科的去、现状和将来的综述；也可是横向的综合，如对各产业部门同类产品的比较综述等。

(2) 针对性：三次文献是为了特定的目的，搜集大量相关的文献，进行分析、综合而编写出来的，并且在很多情况下，它是信息部门接受用户的委托后，进行信息研究后产生的成果，因此，具有很强的针对性，即针对特定用户的信息需求，为特定的目的服务。

(3) 科学性：三次文献是在已有的知识成果的基础上，对特定专业课题的总结和综述，因此，其观点比较成熟，内容比较可靠，有材料、有事实、有数据、有建议、有结论，具有较高的科学性，一般可直接提供参考、借鉴和使用，因而普遍为科研人员和管理者所重视。

(三) 文献信息的主要类型

按文献的性质、特点和编辑出版形式的不同，科学文献又可分为以下十种类型：

1. 图书。大多是对已发表的成果和经验，或某一知识领域的系统论述或总结。它往往以期刊论文、会议论文、研究报告等一次文献为基本素材，经作者的分析、归纳、重新组织而成。图书提供的知识一般比较系统、全面、可靠，起着综合、积累和传递知识的重要作用。从时间上看，图书报道的知识比期刊和特种文献晚，且出版周期较长。

2. 期刊论文。指采用统一名称，定期或不定期出版的连续出版物。其特点是出版周期短，报道速度快，数量大、内容多，发行面广。期刊的内容一般是一次文献，即原始文献，许多新的研究成果大多首先在期刊上发表，因此，期刊论文是获取新信息的主要来源，是极其重要的信息源。

3. 科技报告。是关于某项科学研究和革新成果的报告或研究过程中的阶段进展情况的实际记录。它反映的科学研究和技术革新成果比期刊论文快，内容高度专门化，且具有一定的保密性。一般以单行本的形式出版。

4. 专利文献。专利是国家对发明创造的法律保护。广义的专利文献应该是一切与专利有关的文献，包括专利说明书、专利公告、专利分类表、专利文摘等。狭义的专利文献一般指专利说明书，它是专利文献的主体。专利文献是极其重要的信息源，蕴藏着丰富的科技信息。

5. 会议文献。一般是各种科技会议上的论文或书面发言。随着科学技术的发展，各种科技会议日益增多，已成为科技交流的一种重要渠道。许多学科中的重要发现有很大一部分是在会议文献中公开的，加上一些会议论文不在其它刊物或出版物上发表，或者即使发表，也要经过较长一段时间，因此，会议文献越来越受到科技界的重视，它已成为科技人员了解本专业的发展水平和最新研究成果的有效工具。

6. 政府出版物。指各国政府部门及所属机构发表、出版的文件，大体上可分为行政性文件（如法令、方针政策、规章制度、统计资料等）和科技文献两大类。科技文献中包括政府各部门的科研报告、技术政策等，对了解某一国家的科学技术和经济政策及其演变等情况有一定的参考价值。

7. 学位论文。是为了取得某一级学位而提交的论文，包括学士论文、硕士论文和博士论文。它们的特点是：论文的水平和质量差别较大，论题比较专一，阐述系统、具体，有一定的独创性观点，且经过一定的审查，故有一定的参考价值。

8. 标准文献。是对工农业产品和工程建设的质量、规格及检验方法等方面做出的技术规定。它是从事生产、建设的共同技术依据，是一种规章性的文献，有一定的法律约束力。标准文献属于三次文献，其反映的信息只能达到当时的技术标准和技术水平。因此，其技术新颖性和及时性逊于专利等文献。

9. 产品样本。是厂商为推销产品而印发的一种宣传性出版物，包括产品的性能、规格、尺寸、重量、构造、用途、使用方法等。产品样本的特点是技术上比较新颖，参数比较可靠，能给人以直观形象。利用它可了解国内外技术水平及有关技术的演变和发展动向，可获得设计、制造、使用中所需的数据和方法，也可以为判断产品的价值提供依据。

10. 科技档案。是在科研生产活动中形成的有一定具体工程对象的技术文件、图样、照片、原始记录的原本或复制本，内容包括任务书、技术指标、



研究方案、实验记录、设计图纸等。它是科技活动的真实记录，内容一般真实、准确可靠。

## 第二章 信息检索基础

### 第一节 信息检索概述

#### 一、信息检索的概念

信息检索是信息工作的一项重要内容，是沟通广大信息用户和各种信息源的主要渠道，也是大量情报信息进行广泛传递的集中表现形式，最能体现信息存贮的目的。

##### （一）广义的信息检索

从广义上理解，信息检索是指将信息按一定方式组织和贮存起来，并根据信息用户的需要找出有关信息的过程。所以它通常被称为“信息存贮与检索”，包括各种文献信息的加工存贮和检索利用两方面的内容，是由两个方向相反而又相互依存的工作过程构成。广义的信息检索概念是对信息工作者而言的。

##### （二）狭义的信息检索

狭义的信息检索是指广义的信息检索的后半个过程，即从信息集合中找出所需要信息的过程，相当于人们所说的信息查检等。在实际工作中，也往往把日常的信息加工和存贮视为内部准备工作，而狭义地把信息的查检视为信息检索。狭义的信息检索的概念通常是针对用户而言的，即根据用户的特定要求查找所需信息的全过程。

作为检索对象和目的的信息，其出现形式是不同的，有的以记录着信息或知识的文献表现出来，有的以含着情报或知识的数据或事实表现出来，于是根据检索对象，即信息出现形式，信息检索又可区分为不同的类型。

##### （三）信息检索的基本类型

1. 文献检索。以文献或其索引、文摘为检索对象，目的是核实已知文献的不确切项目，如著者名、年代、出处等，或查检某课题原始文献的线索及它们的具有收藏情况。这类检索数量大，方式不一，是信息检索的基础部分。一般说来，文献检索是一种相关性检索，它只提供有关的文献供参考，而不直接回答用户所提技术问题的本身。例如：“某作家在60年代的著述情况”，就是要根据查寻要求，按照一定的检索标识，从所收藏的文献中查找出特定的文献，以供参考。

2. 数据检索。以数据为检索对象，可直接选择专类的数据性工具，从中查检，从而得到数值性数据、图表、化学结构式、计算公式等等，它们都是经过人们精心测试、绘制、评价、处理而得出来的确切数据。数据检索是为了满足科技工作者对浓缩信息的特殊需求而出现的，这种浓缩的信息，用户可直接使用，无需查阅原始文献，因此可大大节约研究人员的时间，提高工作效率。数据检索是一种确定性检索，要直接回答用户所提问题的本身，提供所需的确切的数据。

3. 事实检索。是以特定的事实为检索对象，先选择适宜的检索工具，按一定检索标识，直接从中检出事实性、知识性的答案。事实检索也是一种确定性检索，要求提供用户所需的确定的事实。

#### 二、信息检索的产生和发展

信息检索最重要的早期发展是文献检索。文献检索是随着文献资料累积到一定程度时才出现的，目的是为了满足不同用户特定的文献查寻和文献需求。

可以说，有了文字记载以后，就有了文献检索的萌芽。但在人类文明发展的早期，从事科学、文化活动的人数不多，文献数量不大，人们查找文献仅仅通过私人书信或其它简易方式，就能达到互通信息的目的，所以没有而且也不必开展信息检索，更无必要建立完备的检索工具。

直到进入 20 世纪以后，由于科学技术的飞速发展，现代记录下来的知识急剧增长，文献数量也迅速增加，从而加大了文献查找的难度，于是真正意义上的信息检索产生了。一些专门的检索工具，如文摘、索引、目录、百科全书等的编纂也随之发展起来，此时，人们查找的着眼点更多地指向包容在文献中的信息而不仅仅局限于文献本身。之后，由于电子计算机在信息工作中的应用，使得人们拥有强有力的信息存贮和信息处理手段，从而促进了信息检索走上了计算机化的道路。

总的来说，信息检索经历了两个主要阶段，即传统的手工检索阶段和现代计算机信息检索阶段。

### （一）手工检索阶段

文献信息检索直接发源于文摘索引工作和图书信息部门的参考咨询工作。

早在西汉时期，刘向、刘歆父子整理编撰《别录》和《七略》，成为最早带有内容摘要的图书目录，开辟了从图书目录直接了解和查找西汉之前书籍概况的先河，是最早的书目性工具书之一。

1665 年 1 月 5 日，法兰西科学院在巴黎创办了《学者周刊》。该刊除报导科学领域的重要事件、科学院的活动和教会法庭的重要决定外，还报导、摘录或评论新出版的图书，成为世界上最早的科学期刊之一，也是以专栏或附录形式出现的最早的文摘刊物。在这以后的一百多年中，许多综合性的、专业性的文摘刊物相继出现，成为一种常用的信息传递方式和检索媒介。

索引工作也有较长的历史。在我国唐宋时代，一些文人学者就编制了一些工具书，供查找古籍中的俚句骈语、诗赋文章、史实或其它资料，人们通常称之为“类书”，实际上它们就是属于索引这一类的工具书。在西方，第一部专门的索引约出现在七八世纪，是为《圣经》编的《圣经语词索引》。之后，西方出版的一些图书开始编有书后索引，以供读者查找书中的特定材料。

到 19 世纪初，文摘刊物开始走向独立编辑出版，而且报刊索引工作也随着报刊文献的增多而得到了很大的发展，并且与文摘刊物紧密结合在一起，成为查找科学文献的最重要的手工检索工具。

图书馆及其馆藏文献的急剧增多，使图书馆的馆藏目录工作迅速开展起来，成为查寻馆藏文献的有力工具。另外，图书馆的参考咨询工作也包括了为读者提供检索服务的任务。

随着人类科研活动的增加，文献种类的增多，人们对文献检索的需求越来越普遍，从而促进了检索工作逐步走向正规化和专门化，检索刊物体系也逐渐形成，检索工具书更趋完善，成为图书信息部门做好文献信息检索的有力保障，信息检索也成为信息工作的重要内容。

### （二）计算机信息检索

20 世纪四五十年代，由于计算机的成功制造和使用，使得文献信息检索逐步迈向机械化和自动化。

1954 年，美国海军兵器中心首先在 IBM701 型电子计算机上成功地建立

了世界上第一个计算机文献检索系统。它将文献号和少量标引词存储在计算机内，匹配命中后输出的是文献号。

1957年，H.P.卢恩等人开始研究采用计算机编制索引取得成功。这一成就激励了图书信息界，许多文摘索引机构相继开展用计算机编制文摘索引刊物的试验。

1964年，美国国立医学图书馆的医学文献分析与检索系统（MEDLARS）建成并投入使用，标志着文摘索引刊物的生产实现了机械化，检索服务实现了计算机化。不久，像美国化学文摘社、工程索引公司等一批著名的文摘机构也实现了各自的目标，开始建立机读文献型数据库，发行磁带版的水文摘索引刊物，供图书情报单位开展文献检索服务。到60年代末70年代初，随着分时计算机的出现和通讯网络的发展，又出现了功能更强的联机检索系统，如现在普遍使用的MED-LINE系统、DIALOG系统以及ORBIT系统等，联机检索服务也迅速地在世界各地得到发展。

计算机技术、通讯技术和文献信息检索的成功结合，使信息检索走向了计算机检索时代。文摘索引刊物的编辑排版工作从落后的手工方式过渡到先进的机械化自动方式，大大地缩短了编辑出版时间，加快了信息的传递和利用。机读数据库的出现和计算机检索的实现，为用户检索信息提供了快速便利的途径。

信息检索将逐步摆脱传统的手工检索模式，而走向全新的、面向全社会的自动化和网络化阶段。

### 三、信息检索的意义

信息检索是在信息用户与信息源之间充当媒介作用，它是联系信息生产者与信息需求者的中间环节，是信息交流和传递的重要过程，是提高文献利用率和科研效率的重要手段。概括起来，信息检索的意义主要体现在以下几个方面。

1. 信息检索是获取知识的有效途径。在当今文献和知识急剧增长的信息时代，如何在浩如烟海的文献信息流中迅速找到所需文献信息，并加以合理有效的利用，是摆在科研工作者面前的一道重要课题。据测算，人类知识总量在19世纪每50年增加一倍，20世纪初每10年增加一倍，20世纪70年代每5年增加一倍，80年代几乎是每3年增加一倍。而且，文献信息的分布已极不平衡，同一学种的论文分散在几十种、甚至几百种期刊上，导致文献的查找越来越困难。因此，如何以最少的精力、最短的时间充分占有文献信息，成了人们亟待解决的实际问题，而信息检索正是有效解决这一问题的最好途径，它可以帮助人们快、准、全地获取所需信息，最大限度地节省查找时间，使文献信息得以充分的利用。

2. 信息检索是科研工作的组成部分。科学研究是一种创造性的思维活动，是在继承前人成果和借鉴他人工作的基础上发展起来的。科学研究一般包括资料准备阶段和研究阶段，其中，资料准备阶段要占用科研人员全部科学研究的30%~60%的时间，还会随着信息检索效率的高低发生变化。高效的信息检索不仅能使科研人员减少查找资料的时间，而且还能够通过信息检索获得启发，对研究工作具有极大的帮助作用，最终缩短科学研究的时间。

3. 信息检索能够避免重复研究和走弯路。科研的任务是为了创造新的知识成果，它要求对某一课题或某一领域的认识及判断应是前所未有的，因此，科研主题必须建立在完全的充分的信息检索基础之上，否则很容易造成科研

项目的重复研究，从而造成人力和物力的严重浪费。据统计，我国的低水平的重复研究现象比较严重，尤其是省级及以下科研项目，重复率达到 50%。可见，要进行有价值的科学研究，必须全面地获取有关文献信息，及时了解各学科领域出现的新问题、新观点，这只能依赖文献信息检索才能实现。通过信息检索，可以掌握研究课题的历史和现状，了解其发展趋势，以确定自己的研究起点和研究目标。

## 第二节 信息检索的原理和方法

### 一、信息检索的基本原理

如前所述，信息检索的全过程包括了信息的存贮和信息的检索两个过程。其中，存贮是为了检索，而检索必须先要存贮，信息存贮和信息检索是方向相反而又相互依存的两个方面。

在信息工作中，检索可以理解为“检”与“索”的统称，“检”是“选检”、“归类”，指从众多的加工对象中挑出性质相同的集中起来，经筛选、整理、分析、标引，将得来的信息产品归纳到一个统一的系统中，即从众多的信息源中筛选合乎既定要求的情报信息，予以标引、存贮和积累，以待利用。“索”即“索取”，是“存入”的逆过程。所以，信息检索是一个获得情报信息的操作过程，其实现必须借助于一个特定的信息系统，而且检索的出现也必须以存贮的存在为前提条件。如无大量的存贮，则根本谈不上从中选择和提供信息。

同时，信息检索要求存贮过程和检索过程必须采用相同的检索标识进行信息标引，使要存入的信息特征与要检索的具有一致的特征标识形式。

信息检索的基本原理就是检索者将检索提问的标识与存贮在检索工具中的信息特征标识进行比较，结果凡是信息特征标识与检索提问标识相一致，或者信息特征标识包含了检索提问标识，那么，具有这些特征标识的信息就从检索工具中输出，输出的信息线索与检索者所需的信息线索大致吻合。

从一定的意义上说，信息检索成败的关键，是能否用规定的检索标识系统（即检索语言）来正确标引检索提问，而正确标引检索提问的关键又在于是否能从检索工具的词表中选出最能确切表达检索提问所需要的标识。

### 二、信息检索方法

信息检索的效率与具体的信息检索方法有很大的关系，运用有效的信息检索将能够以最少的时间获得最满意的检索效果。归纳起来，信息检索方法主要有以下三类。

#### （一）手工检索方法

这是查找印刷型文献信息的传统方法，一般分为以下四种。

1. 直接查检法。指不依靠检索工具，而通过浏览或查阅原始文献直接获取文献信息的方法。其优点在于，能够明确判断文献所包括的信息是否为自己所需要的信息，缺点是难以获得全面的文献，且费时费力。如果检索课题单一，文献相对集中，又熟悉检索书刊，则可用这种检索方法，而对有多个主题、文献离散度较大的课题，就难以获得理想的检索效果。由于此方法不依靠检索工具，因此不是严格意义上的文献信息检索方法。

2. 间接检索法。即借助于检索工具获取所需文献的方法，一般包括顺查法、倒查法和抽查法。

(1) 顺查法：是一种从旧到新的顺时序的查检方法，一般需要了解检索课题的背景、发生和历史简况，再通过有关的参考工具核实和深入了解该课题的实质性内容和概貌，从而选择比较适宜的检索工具，从问题产生的年份着手查起，直到满意为止。开始选材时可适当放宽范围或要求放松一些，待发现这类信息源相当丰富时，可缩小范围或要求严一些，以避免漏选而返工重检。但逐年的查检，劳动量因覆盖面大而随之增大，检索效率不高，多在缺少综述性文献时采取。其优点是查全率较高。

(2) 倒查法：指由新而旧的逆时序的查检方法。此法多用于查找新课题或用于为老课题查找新资料。课题对近期的状况比较重视，从新情况开始查到一定的基本资料时为止，时间终点视课题要求而定。其优点是省时省力，检索效率较高，但查找资料不如顺查法齐全，容易出现漏检，因而对课题研究的全貌不易把握。

(3) 抽查法：根据课题的要求，针对所属学科处于发展兴旺时期的若干年进行文献信息查找。用这种方法能获得相对集中、具有代表性、且能反映学科发展水平的文献信息，往往能起到事半功倍的效果。其优点是检索效率高，检索效果好，但要求在检索之前须掌握该学科的发展情况，熟识该项技术发展的特点，以便正确地选择抽查的时间范围。

3. 追溯法：也称为文献追踪法。此法不是利用确定的检索工具，而是利用已知文献的某种指引，如文献附的参考文献、有关注释、辅助索引、附录等，追踪查找文献。根据已知的文献指引，查找到一批相关文献；再根据相关文献的有关指引，扩大并发现新的线索，去进一步查找。如此反复追踪扩展下去，直到检索到切题的文献。用追溯法检索文献，最好利用与研究课题相关的专著与综述，因为它们所附的参考资料既多且精。此种方法一般在缺乏检索工具或对检索工具的使用不熟悉，以及文献线索很少的情况下使用。其优点是简单方便，容易查找。缺点是漏检和误检的可能性较大。

4. 综合法：是以上各种检索方法的综合使用，即先通过选出的检索工具查出一批相关文献，然而再利用这批文献所附的参考文献进行追溯查找，从而得到更多的相关文献，如此交替循环使用，直至满足检索需求为止。综合法是一种“立体型”的检索方法，其检索效果较好。

## (二) 机械检索方法

指借助一定的机械进行检索的方法。所用机械如卡片检索机、缩微胶卷检索机、电子计算机等。当前，一般说来，机械检索就是指电子计算机检索。

电子计算机检索具有检索速度快、检索途径多、检索效果好等特点，可人机对话随时变化检索要求，并可显示阅读或联机、脱机打印和输出符合检索需求的部门。1964年，美国国立医学图书馆建立了医学文献分析与检索系统，并于1971年发展成联机检索系统，目前已成为世界上应用最广泛的计算机医学文献检索系统，其检索功能也越来越强，从仅能用主题词检索发展到可以用自由词、字符串、词干、词头以至句子、符号、公式等检索。计算机检索已成为信息检索的必然发展趋势。当前，光盘的出现及其与微型电子计算机的联合应用，使信息检索又发展到一个新的时期。光盘检索系统建立方便，不受检索时间及通讯线路的限制，能够深入到千家万户，已成为电子计算机信息检索普及应用的较为理想的检索系统。另外，随着计算机网络进一步发展，计算机网络化信息化将成为新的趋势。

电子计算机检索的方法是以概念组配系统为基础的概念组配方法、检索

时通常需把用户信息提问的复杂概念分解为若干单元概念,此时,各个单元概念仅能表达用户需求的各个侧面,必须将各个单元概念加以恰当的逻辑组配,才能表达出用户提问的一个完整概念内容,即表达单元概念的检索词加以恰当组配,形成检索式,才能检索出所需的文献信息。

电子计算机信息检索在西方发达国家已得到普遍应用,在我国大中城市也有一定的普及,而一些中小城市由于条件所限,其应用范围还比较窄,还未得到推广。

### (三) 非正式方法

除传统的手工检索方法和先进的电子计算机信息检索方法外,还可以利用另外一种获取信息的重要方法——非正式方法。

非正式方法通常也称之为非文献方法,是指不借助文献而获取信息的方法。它是一种较为古老的信息获取方法,在印刷术发明前,它是交流、获取信息的主要方法,目前已不起主导作用,但仍很重要。非正式方法主要通过交谈、书信往来、参观访问、考察、参加学术会议、听演讲、实物样品的搜集及技术考察等方式获取所需的信息,一般比从正式发表的文献中获取的信息要早得多,快得多。据日本学者报道,在学术杂志上发表的论文,约 1/5 已在学会内部刊物刊登,约 1/10 已在地方学术会议上宣读,约 60% 论文作者在原稿完成后即已复制送给同行阅读,论文正式出版前,以口头或文章形式将其内容作为信息传递的达 67%。可见,非正式获取信息的方法不失为一种较好的方式,具有迅速性、新颖性、针对性和直观性的优点,其缺点是难以评估信息的价值,客观性和真实性也难以保证,且适用范围有限。

## 第三节 信息检索的途径和程序

### 一、信息检索途径

检索途径依赖于文献信息的特征。文献具有两种特征:即外部特征和内容特征。文献的外部特征主要是指文献载体上 标明、易见的项目,有文献题名、责任者、序号、出版者、出版地、出版年等;文献的内容特征有所属学科及所属主题等。因此,根据文献的外部特征和内容特征,将信息的检索途径分为两大类型。

#### (一) 文献外部特征的检索途径

1. 责任者途径。即通常所说的著者姓名途径。责任者是指对文献内容负责或作出主要贡献的个人或团体,包括著者名、评者、编者等。责任者途径是根据文献著(译、编)者的名称查找文献信息的途径,是外文检索工具较为重要的途径和惯用途径。按著者姓名字顺排列,易于利用,又便于编排,也易于机械加工。

使用著者途径检索文献信息须注意文种不同和姓名排列方式的差异,如单姓、复姓、父母姓连写、本名、教名以及姓名中附加荣誉称号等。欧美人的姓名习惯上名在前、姓在后,而目前使用的各种著者目录和著名索引则按姓在前、名在后的方式以字序排列,因此,在具体检索时应按姓在前、名在后的字顺查找。

2. 题名途径。也称书名途径。题名是表达、象征、隐喻文献内容及特征的词或短语,是文献的标题或名称,包括书名、刊名、篇名等。文献题名有正题名、副题名和辅助题名。题名检索途径是指根据文献题名查找文献信息

的途径。它把文献题名按照字顺排列起来编成索引，其排法简单易行，易于查检。但因书名和篇名较长，不宜作为检索标识，又因不同文字的形体结构和语法结构有自己的特色，字尾变化复杂，所以难以把同样意义的文献集中于一处，实际使用价值已不为人们看好，而逐渐失去重视。

3. 文献类型途径。文献信息检索工具收选的信息源多种多样，如期刊、图书、科技报告、专利、技术标准、政府出版物、会议录等。为满足查询者不同的检索要求，如会议文献或专利文献的查找，不少检索工具也增设文献类型检索途径，如专利号索引、图书索引、会议索引、报告号索引等，以满足不同类型检索用户的需求。

## （二）文献内容特征的检索途径

1. 分类途径。是指按文献内容的学科分类体系查找文献信息的途径。一般说来，一种检索工具的编制都须按学科建立自己的分类体系，其收录的文献按分类目录中的排序进行编排，这样编排的结果可将同一学科的文献集中，便于按学科查找文献。分类目录和分类索引是普遍使用的分类检索工具。其缺点是，对于较难分类的新兴学科和边缘学科来说，查找不便，查找时须首先了解反映学科体系的分类表，再将概念变换为分类号，然后按分类号进行检索，由于概念变换为分类号的过程中易出差错，所以也会导致漏检和误检。

但是很多检索者希望从其熟悉的分类系统，从学科概念的上下左右关系了解事物的派生、隶属、平行等关系，满足族性检索的需求。分类途径能够较好地满足这一要求。

2. 主题途径。主题是文献所表达的中心思想、所讨论的基本问题和研究对象。主题途径指根据表达文献主题内容的主题词及其派生出的关键词为标识查找文献信息的途径。其主要检索工具是主题目录和主题索引，或标题词索引、关键词索引、叙词索引等。主题目录按文献内容主题词组织，以文献所讨论的主题直接检索，可以查到分散于各学科里同一主题的文献。主题索引是工具书辅助索引之一，它可揭示包含该主题的文献信息在文献正文中的位置。

主题途径检索文献信息的优点是，用主题词作为标识，表达概念准确、灵活、专指度高，可使同一主题的文献集中，检索效率高。又由于主题词可随科技发展增加或更新，因此便于查找新兴学科的文献信息，在各学科和其分支交叉渗透日益增多的当前，主题途径较好地适应了这一要求。但其缺点是，主题索引缺少学科系统的整体性和层次性，因此，难以达到很高的查全率。

3. 分类主题途径。是分类途径与主题途径的结合，它能够尽量避免两者的不足，取其所长。一般说它比分类体系更具体一些，无明显的学术层次划分，又比主题法概括一些，但保留了主题体系按字顺排序以便准确查检的特点。

4. 其它途径。主要是指一些辅助检索途径，如按化学分子式排出的“分子式索引”，可提供一种从分子式角度查找化学化工文献的目的，另外还有化学物质索引、合金索引、地名索引等各种专门索引，以满足查检特定种类文献信息的需求。

## 二、信息检索的程序

信息检索工作是一项实践性和经验性很强的工作。对于不同的待检课



题，将采用不同的检索程序，即文献信息检索的具体步骤和方法应因题而定，因人而异。但在实际检索工作中，还是可以依据信息检索的基本原理，归纳出文献信息检索的一般程序和步骤，以使检索工作有条不紊，取得较好的检索效果。

通常，信息检索接下列程序进行：

#### （一）分析检索课题，明确提问要求

所谓检索课题是根据查找文献信息或查解科研疑问的需要所拟定的问题。在检索之前，首先须对待检课题进行认真分析，明确检索的目的和具体要求，以便使用户的检索提问符合检索工作的要求，做到检索与提问一致，避免盲目检索。

在分析课题的基础上，还要辨明检索课题的类型，是查文献，或是查事实，还是查数据，以及要求查找文献信息的时间范围、学科范围等，以求对检索课题有个总的认识。

#### （二）选择检索工具

当检索课题明确以后，就要选择与课题相符、质量较高、检索手段比较完善的检索工具。主要考虑该领域有哪些适用的工具，其出版时间和所概括的范围能否包括拟检的问题，其质量和权威性如何，是否具备多种检索途径等。

一般对检索工具的要求是：收录文献资料的专业广、类型齐全、数量大、报道速度快、文摘详细，并附有各种索引。检索时，既要选择专业性检索工具，也要考虑使用综合性检索工具，以获得满意的查全率。

选择检索工具时，可利用有关的指南、检索手册及便览等，以对各种工具书进行比较和选择，从而确定更适合课题要求的检索工具。

#### （三）确定检索途径和检索方法

检索途径和检索方法的选择，取决于检索课题的要求和已掌握的情况。如果课题检索的泛指性较强，即所需文献的范围较广，则选用分类途径较好，检索方法可以选用顺查法，以全面了解课题的研究背景、发展及现状；反之，若课题检索的专指性较强，即所需资料比较专深，则选用主题途径为好，检索方法可选用倒查法和抽查法，以获得该课题的最新文献信息。检索途径和方法选择不当，将造成误检和漏检，从而影响检索效果。

#### （四）实施信息检索

以上所述皆为检索的准备阶段，有些更带有策略性的，在检索前应进行全面的考虑。检索过程的实施一般可分三步进行。试查：按已选定的检索工具和方法，抽样或小范围初查一下，若发现问题，可对检索方法作适当修改。正式查：这是信息检索的主体，主要是利用检索工具进行查找，通常利用工具的目次、分类表、主题词表、类目索引或检索手册等辅助性工具，以使检索更为直接准确。补查：检查结果发现仍不能满足需要，或发现新的查找线索，可进一步做些补查工作。

#### （五）获取原始文献

根据检索所得的题录或文摘，可以查找到原始文献的作者、所在刊物及收藏单位，然后利用馆藏目录或其他各种目录，以馆藏中查找原文或向其他收藏单位索取原文或复印件，也可通过查得的作者联系地址或作者单位向作者本人索取。至此，一次信息检索的过程终止。

#### 第四节 信息检索效果的评价

所谓检索效果，就是利用检索系统（或检索工具）开展检索服务时产生的有效结果。检索效果评价着眼于整个检索系统，着眼于系统的使用效果和服务质量，因而它直接反映了检索系统性能。

目前，衡量检索效果的方法主要有三个方面：检索结果有效性的评价，即检索的技术效果的评价，这是对检索系统检出相关文献以满足用户提问要求能力的一种测度，主要以查全率和查准率为评价标准。检索系统实用性的评价，包括系统对用户是否需要，是否实用，有多大的实用效果，即检索的社会效果的评价，其中要涉及到社会学及其方法。检索费用——效率评价，即检索的经济效果的评价，包括检索服务的成本和时间消耗，这涉及到信息检索系统的经济学问题。然而，由于检索效果的评价涉及到许多问题，可以从不同的角度采用不同的检索效果评价方法。其中，最常用的检索效果评价指标为查全率和查准率。

##### 一、查全率与查准率

查全率和查准率是由 J.W. 佩里和 A. 肯特于 50 年代中期提出来的，后经不断改进和完善，至今已成为评价检索效果最常用的两项关键指标。

确定查全率和查准率最常用的方法是有名的  $2 \times 2$  表。

表  $2 \times 2$  表

用户相关性判断	相关文献	非相关文献	总计
系统相关性预报 被检出文献	a	b	a+b
未检出文献	c	d	c+d
总计	a+c	b+d	a+b+c+d

它反映检索系统在实施某一次检索时所得的结果

其中 a 表示被检出的相关文献，即查准的文献；

b 表示被检出的非相关文献，即误检的文献；

c 表示未检出的相关文献，即漏检的文献；

d 表示未检出的非相关文献，即正确拒绝的无关文献。

存贮在文献检索系统文档中，参加检索的全部文献量为  $(a+b+c+d)$ 。从检索系统角度来看它们可以分为两部分：一部分是被检出文献  $(a+b)$ ，也就是与检索策略相匹配的部分；另一部分是未检出文献  $(c+d)$ ，即与检索策略不相匹配的部分。因为两部分文献反映了检索系统处理是否与检索提问相关，故称之为系统相关性预报。

另一方面，从用户的角度来看，检索系统文档中参加检索的全部文献也可以分为两个部分：一部分与用户需要相符，称为相关文献  $(a+b)$ ；另一部分与用户需求不符，称为非相关文献  $(b+d)$ 。因为这两部分文献反映了用户判断是否与检索需要相关，故又称之为用户相关性判断。

这样，便可以根据  $2 \times 2$  表来确定查全率与查准率的含义。

查全率就是系统在进行某一检索时，被检出的相关文献量与系统文档中实有的相关文献量的比率，可用下式表示：

$$\begin{aligned}\text{查全率 (R)} &= \frac{\text{被检出相关文献量}}{\text{文档中相关文献总量}} \times 100\% \\ &= \frac{a}{a+c} \cdot 100\%\end{aligned}$$

查全率就是被检出的相关文献量与被检出的文献总量的比率，可用下式表示：

$$\begin{aligned}\text{查准率 (P)} &= \frac{\text{被检出相关文献量}}{\text{被检出文献总量}} \times 100\% \\ &= \frac{a}{a+b} \cdot 100\%\end{aligned}$$

可见，查全率是用来描述检索系统检出相关文献的能力；查准率用来描述检索系统拒绝非相关文献的能力。

在具体评价检索系统的检出效果时，一般应将查全率与查准率结合起来，否则难以准确反映检索系统的功能和检索效果。

虽然查全率与查准率能较好地反映出—个检索系统的检索效果，但是，它们也存在一些难以克服的局限性和固有的缺陷。首先，—个检索系统中总共有多少相关文献 (a+c) 难以确切计算，而只能是大概估算；其次，对于全部相关文献对用户的价值是建立在假定具有同等价值上的，然而实际上并非如此。不同的用户对相关文献的认识也可能不一致，因此存在着太多的主观成分和一些模糊概念，所以说明上述方法求得的查全率与查准率并不是绝对的，而只能是相对近似地描述检索效果。

分别与查全率和查准率相对应的指标是漏检率和误检率，可以用下面的公式加以表示：

$$\begin{aligned}\text{漏检率 (M)} &= \frac{\text{未检出的相关文献}}{\text{文档中相关文献总量}} \times 100\% \\ &= \frac{c}{a+c} \cdot 100\% \\ \text{误检率 (N)} &= \frac{\text{检出的不相关文献量}}{\text{检出的文献总量}} \times 100\% \\ &= \frac{b}{a+b} \cdot 100\%\end{aligned}$$

从上述式子可以看出：R+M=1，R+N=1。

此外，还可以求出离散率和正确拒绝率的公式：

$$\text{离散率 (F)} = \frac{b}{b+d} \cdot 100\% , \text{它表示不该检出而被检出的文献量的比率。}$$

$$\text{正确拒绝率 (Rs)} = \frac{d}{b+d} \cdot 100\% , \text{它表示正确拒绝检出无关文献量的比率。}$$

## 二、影响查全率与查准率的因素

从查全率与查准率的定义可以看出，这两个指标主要是面向用户的指标。用户是一方，由检索人员与检索工具组成的检索系统是另一方，对于用户提出的每一查找要求，检索系统作出相应的响应，这两个指标反映了用户对检索系统响应的主观评价。因此，影响查全率与查准率的各种因素主要来自用户与检索人员的配合、检索策略、标引和检索语言等方面。

1. 标引的影响。标引人员标引文献的正确性对查全率与查准率有直接影响。标引人员在标引文献时主要面临两方面的问题：确定文献中含有哪些有用信息及它们能够响应哪些查找要求；如何用给定的检索语言表达这些检索要求。其中第一个问题更关键些。如果提取出来的不是有用信息，则以后查找时查准率降低；如果有用信息没有被提取出来，则以后查找时查全率降低。有用信息提取出来之后，获得正确标引的主要困难在于确定需要这些有用信息的用户会提出怎样的查找要求，如果将能够响应的某个查找要求遗漏掉，则用户从这个查找中就查不到会有该有用信息的文献，因而查全率会降低。

文献所能响应的查找要求确定出来之后，在将其转换成标引词的过程中，标引人员可能发生两种类型的差错：遗漏了必须描述的概念，从而使查全率降低。发生这种差错的原因是标引人员的疏忽，或者因检索语言中没有相应的标引词，也没有相应的使用参照，标引人员不知如何标引而略去。

选用了不合适的标引词，使得检索人员利用这个词查出的文献是无关的，因而引起查准率降低；或者检索人员利用正确的标引词查找时，有关文献因标引了不合适的标引词而查不出来，则导致查全率降低。

此外，标引的网罗性与专指性对查全率与查准率也有影响。标引的网罗性高，查全率也高，但查准率低；标引的专指性高，查准率也高，查全率则较低。

2. 检索语言的影响。由检索语言引起的查找失败有两种类型：一是因标引词专指性不足引起的查找失败；一是因标引词之间的含糊关系或虚假关系引起的查找失败。

标引词的专指性不足既会造成查全率降低，又会造成查准率降低。

标引词的专指性是影响查准率的积极因素，但高专指性的标引词是提高查准率的同时有降低查全率的趋向。这是因为标引词的数目越多，能够表达的意义差别也就越细致，标引就越难于取得一致。

如果检索语言中的标引词不规范，或允许使用过多的标引词，则可能引起虚假组配现象，将使查准率降低。

词表结构对查找有很大影响。如果词表不以某种方式把所有关连的标引词集中在一起，那些检索人员就不能将与查找要求有关的全部标引词找出来，查全率就会降低。词表对标引也有很大影响，较好的词表参照系统和等级结构关系能够提高查全率。

3. 检索策略的优劣。所谓检索策略是指检索者对检索的每一步骤所做的安排和部署，如选择检索工具、检索方法、检索途径等。它主要取决于检索人员的知识水平与业务能力，因此，检索策略的优劣是影响检索效率的主观原因。

4. 查全率与查准率的互逆相关性。英国 Cleverdon 等人由试验得到的查全率—查准率经验曲线表现，查全率与查准率是互逆的，即对于一个查找要求，如果要设法提高查全率，则查准率会降低，反之亦然。这也可以看作是文献信息检索的一个基本特征。

### 三、提高检索效果的措施

一般地说，检索人员与用户对于检索效果的要求是一致的，既要求有较高的查全率，又要保证有理想的查准率。然而，具体到每一个用户，则他们对检索效果的要求可能不一样，这取决于他们检索的目的。如果用户仅仅想

了解某一专业领域的发展概况，则对查全率有较高的要求；如果用户检索是为了解决工作中的某个技术难题，就需要较高的查准率。因此，检索人员可以根据实际用户需要合理地调节查全率和查准率。

总体来说，提高检索效果可以采取以下措施：

1. 提高检索工具或检索系统的编辑质量。这就要求检索工具的收录范围要全面，内容要准确、详尽。同时，要求检索工具提供不同途径的检索入口，即辅助性索引要齐全。2. 提高索引语言的专指性，加强对索引词汇的控制，完善词表的结构及其参照关系，使索引语言更有利于族性检索，也有利于特性检索，同时兼顾查全率和查准率。

3. 提高标引质量。标引人员进行标引时，应做到不错标、不漏标、不滥标，使标引词的定义和使用范围与检索系统的文献内容相一致，从而提高检索的准确性。

4. 提高检索人员的业务水平，力争做到检索前正确理解检索课题的实质要求，制订最优的检索策略，同时，在检索过程中保持认真负责的精神，减少人为的错检和漏检。

## 第三章 信息检索语言

### 第一节 信息检索语言概述

#### 一、信息检索语言的定义

信息检索语言是根据信息检索的需要而创造的人工语言。一般又称为信息、存贮与检索语言，索引语言，标引语言，文献工作语言，标识系统等。其实质是用于表达一系列概括文献信息内容的概念及其相互关系的概念标识系统。它可以是从自然语言中精选出来并加以规范化的一套词汇，可以是代表某种分类体系的一套分类号码，也可以是代表某一类事物的某一方面特征的一套代码，用于对文献内容和信息需要进行主题标引、特征描述或逻辑分类。

目前世界上有许多种信息检索语言，比较著名的有：《中国图书馆图书分类法》、《汉语主题词表》、《国际十进分类法》、《杜威十进分类法》、《NASA 叙词表》等，都是信息检索语言的一个语种的。对图书文献资料所标的分类号或检索词，即是信息检索语言的语词。

信息检索语言通常应具备下列三个基本要素：

1. 有一套用于构词的专用字符。字符是检索语词的具体表现形式，它可以是自然语词中的规范化名词或名词性词组，也可以是具有特定含义的一套数码、字母或代码。

2. 有一定数量的基本词汇用来表述各种基本概念。基本词汇是指组成一部分类表或词表中的全部检索语词标识之总汇，如分类号码的集合就是分类语词的词汇，一个标识（分类号、检索词、代码）就是一个语词。而分类表、词表等可以说成是检索语词词典，是把自然语词转换成检索用语的工具。

3. 有一套专用语法规则来表达由各种复杂概念所构成的概念标识系统。标识是对文献信息特征所做的最简洁的表述。标识系统是对全部标识按其一定的逻辑关系编排组合成的有序的整体。语法是指如何创造和运用那些标识来正确表达文献信息内容和信息需要，以有效地实现信息检索的一整套规则。

任何一种信息检索语言，都要采取一定的方法和手段，将它的各种要素按照一定的结构结合成一个有机的统一体，以便发挥其最佳功能。概括起来，信息检索语言的主要功能大致有如下四点：

- (1) 对文献的信息内容（及某些外表特征）加以标引；
- (2) 对内容相同、相关的信息加以集中或揭示其相关性；
- (3) 对大量信息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。

#### 二、信息检索语言在情报检索中的作用

信息检索语言作为信息检索专用语言，它在信息检索过程中所起的作用是极为重要的。

信息检索的全过程包括信息的存贮和信息的检索两个方面。信息存贮是指编制检索工具和建立检索系统。情报检索即是利用这些检索工具和检索系统来查找所需的信息。所以，信息的存贮和检索是两个有着密切联系的过程，其中，信息检索语言在这两个过程中的作用以及这两个过程的相互关系如图 1 所示：

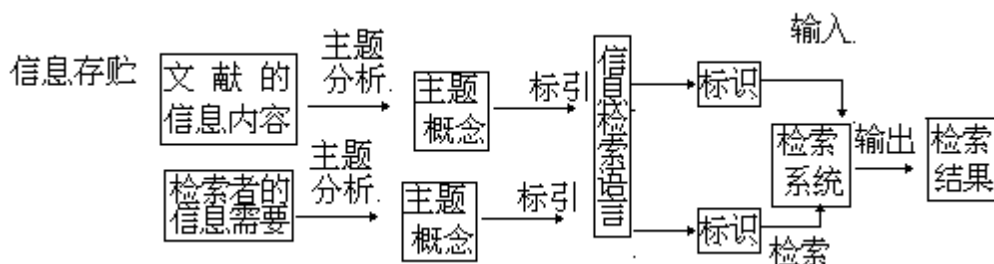


图 1 信息存贮与检索过程及信息检索语言在其中的作用

当存贮信息时，文献标引人员首先要对各种文献进行主题分析，即把它所包含的信息内容分析出来，使之形成若干能代表文献主题的概念，并用信息检索语言的语词（标识）把这些概念标示出来，然后纳入检索工具或检索系统。

当检索信息时，信息检人员首先对检索课题进行主题分析，即把它所涉及的检索范围明确起来，使之形成若干能代表信息需要的概念，并把这些概念转换成信息检索语言的语词（标识），然后从检索工具或检索系统中查找用该语词标引的文献，从而找到包含所需信息的文献。

由此可见，信息检索语言是信息检索系统的重要组成部分，在信息检索系统中起着语言保障的作用，是沟通信息的存贮和检索两个过程，标引人员和检索人员双方思想的桥梁。如果没有信息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信息内容的表达（标引用语）和检索人员对相同内容的信息需要的表达（检索用语）取得一致，信息检索也就不可能顺利实现，甚至根本不能实现。

信息检索语言作为信息检索专用语言的特点，是能简单明白又比较专指地表达文献及检索课题的主题概念，容易将概念进行系统排列，在检索时便于将标引用语和检索用语进行相符性比较；语词与概念一一对应，排除了多词一义、一词多义和词义含糊的现象，并且还能显示出概念之间的相互关系，从而能帮助信息检索人员又全、又准、又快地检索到含有所需信息的文献，可见，信息检索语言对提高信息检索效率，特别是提高查全率和查准率具有重要的作用。

### 三、信息检索语言的基本要求

信息检索语言的质量高低对信息检索效率有重大影响，因此，为了保证信息检索的实施能够达到满意的效果，就必须对信息检索语言提出一些共同的基本要求。

1. 要求信息检索语言保证具有较高的查全率和查准率，力求把信息检索系统中的有关用户需求信息尽量多地查检出来，并使漏检率和误检率控制在允许的范围之内。

2. 必须易于标引和易于检索。信息检索语言不仅要易于理解，易于为标引人员和检索人员所掌握，而且在加工过程中更要易于标引，在检索过程中易于应用。保证易标、易检的因素主要包括：

- (1) 语词或代表符号的含义必须明确而不能含混不清；
- (2) 丰富的语言概念，概念的面广，使遇到的文献信息都能有类可归，有号可用；
- (3) 标识的可观性，使标引人员及检索人员易于识别和理解；
- (4) 查词查号手段多样，便于标引人员和检索人员从各种角度都能方便

地查到所需的标识。

3. 具有对先进的检索方式和检索设备的适应性，以及对多种检索方式和检索设备的适应性。

检索设备可分为传统检索设备和非传统检索设备。传统检索设备如普通卡片式目录、书本式目录；非传统检索设备如机械式检索系统、充电式检索系统、电子计算机检索系统等。

各种检索方式和检索设备都有一定优缺点和适用范围，所以信息检索语言应能适应多种检索方式和检索设备，使其发挥更大的效应。如《汉语主题词表》就兼顾了电子计算机检索系统和手工检索系统两个方面的需要。

4. 能满足多种检索要求。信息检索语言使用的标识多，能概括信息的外部特征和内容特征，满足从不同方面、不同层次进行检索的要求，提供较多的检索途径。

5. 具备与其它信息检索语言的兼容性和一定地域范围内的通用性。每一种信息检索语言都有自己的特点，但又不宜忽视检索语言之间的相互交流及推广使用，这就要求实现各种信息检索语言的兼容性。尤其是要适应当前信息检索自动化和网络化的发展趋势，更应要求信息检索语言向国际标准化靠拢，以求达到更大范围的通用。

## 第二节 检索语言的基础与分类

### 一、检索语言的基础：概念逻辑和知识分类

#### (一) 概念逻辑

检索语言的最终目的是表达一系列概括文献信息内容的概念及其相互关系的概念标识系统。因此，它们都是建立在概念逻辑的基础上的。概念是语言包含的内容，语言是概念的表达形式。事物——概念——语言三者之间——对应的辩证关系既是建立检索语言的基础，也是建立检索语言必须遵循的原则。

概念逻辑，是一种科学思维方法。它通过明确各种概念及其相互关系而揭示事物的本质属性及各种事物之间的联系与区别。检索语言在建立词汇、显示概念关系以及在文献标引和检索中使用，都离不开概念逻辑。所以，它们只有以概念逻辑为基础，才能使检索语言具有广泛的通用性和适应性。

概念是事物本质属性的概括。任何概念都有其内涵与外延。某一概念的内涵，即指及所反映事物的本质属性；外延则是指其反映事物的范围。它们之间存在着“反变关系”，即概念的内涵包含的属性增加，概念的外延便缩小；反之，概念的内涵减少，其外延就扩大。

概念之间的各种关系可归结为相容关系与不相容关系。概念的相容关系包括同一关系、隶属关系、交叉关系、不互相排斥的并列关系；不相容关系包括互相排斥的并列关系、矛盾关系、对立关系。

检索语言在表达各种概念及其相互关系时，普遍地应用了概念逻辑的原理，并且有效地利用了“概念的划分与概括”和“概念的分析与综合”这两种逻辑方法来建立自己的结构体系。

1. 概念的划分与概括（分类），就是利用概念内涵由反映事物本质属性的概念因素构成，概念因素的增加或减少可以形成新的概念，概念内涵与外延成反变关系等的性质，对概念进行划分（缩小）或概括（扩大），形成更



为专指或更为泛指的新概念，用以区别客观世界千差万别的事物；并利用划分或概括过程中所产生的概念隶属关系和并列关系，建立某种形式的检索语言结构体系，即概念等级体系，用以显示客观世界千差万别的事物之间的内在联系。这种结构具有很好的系统性。例如，体系分类法就是应用此种逻辑方法的典型。

2. 概念的分析与综合（组配），是利用在概念的交叉关系中两个概念外延的相同部分可以形成一个新概念，其内涵等于原来两个概念内涵之和，并且它与原来的两个概念具有隶属关系的这种性质，进一步发展为将一个内涵较深的概念分解为两个或两个以上内涵较浅的概念，以及将两个或两个以上内涵较浅的概念合成为一个内涵较深的概念的一种概念逻辑方法，用以建立另一些形式的检索语言结构体系，即概念组配体系。

这种结构体系可以提供从多种途径来进行信息检索的功能，而且可以任意选择检索标识的专指度，根据实际需要扩大、缩小或改变检索的范围。例如，叙词语言与组配分类法便是应用概念分析与综合的典型。

## （二）知识分类

知识分类包括事物分类和学科分类。它将客观世界事物以及各种学科根据当代科学认识的最新成就构成一个完整的、有科学认识意义的体系。所以，它是对千差万别的事物作系统研究的重要方法，是对各种事物之间的区别和联系从本质上、原理上进行揭示的重要手段，对文献信息的系统化具有重大的价值。

为了使信息检索语言能够适合科学研究的需要，必须把各种概念之间的关系建立在知识分类的基础上。如果说概念逻辑是检索语言的基础，则知识分类便是概念逻辑的基础。信息检索语言中的体系分类法、组配分类法、叙词法、标题法以及代码系统等，都在不同程度上反映了知识分类。但是，知识分类体系是多维性和处在不断变化中的，信息检索语言能在多大程度上反映当代知识分类水平，是其质量的重要标志之一。只有较充分反映知识分类的信息检索语言，才能保证信息检索语言在信息检索工作中得到有效运用，并达到较好的检索效果。

### 二、信息检索语言的分类

信息检索语言有很多类型，可以按不同的方式和标准划分。

#### （一）按语言的结构原理划分

1. 分类语言。其又可分为：

（1）体系分类语言：按学科体系层次，从上到下，从综合到一般，逐级展开，属层累制结构。

（2）组配分类语言：一般按学科性质分组，称为“组面”，标引时，选择“组面”和有关词汇加以组配。它是一种新型分类方法，其原理是用有关词汇结合文献信息的内容加以组配。

（3）混合分类语言：是体系分类语言和组配分类语言的结合，又因对二者的侧重之不同又有体系—组配分类语言和组配—体系分类语言之别，但都是先组配。

2. 主题语言。按照主题词性质的不同，又可分为：

（1）标题词语言：从标题词作为文献信息内容的标引标识和检索标识。标题词一般分为主标题和副标题，词表中按级一一列举各级标题，把主标题和副标题加以组配，作为标引和检索使用中的依据。

(2) 叙词语言：以规范化的标引词形成词表，各词独立，标引时根据需要进行组配，可表达文献内容方面的复杂概念，属后组配式，适用于机检和手检。

(3) 关键词语言：是一种无词表的自然语言。关键词选自文献的题目和内容，属于非规范的后组配式。

(4) 单元词语言：以文献中的单元词表达文献信息的内容性质，一般无词表。单元词是文献内容中最基本的不能再加复分的知识单元，单元词之间可以进行组配，属后组配式。

(5) 引文语言：利用文献的引文关系而建立的一种自然语言，无词表，标引词为文献中的主要著录项目，属后组配式。

## (二) 按照语言标识的组配方式划分

1. 先组式检索语言：指表述文献主题概念的标识在检索之前就已固定的检索语言，如体系分类语言、标题词语言均属于这种类型。这种语言标识明确，系统性较好，适用于传统的文献单元方式的目录索引，是检索用户比较习惯的形式。

2. 后组式检索语言：指表述文献主题概念的标识在检索之前未固定组配，而是在检索时根据实际需要按组配规则临时组配的检索语言，如单元词语言、叙词语言均属此类。这种语言采用概念分析和综合的原理，可实行多途径、多因素检索乃至精确检索，相当灵活，检索效率较高。但是，标识明确性不强，造成检索用户使用不习惯。

## 第三节 分类检索语言

分类语言是以号码为基本字符，以科学分类为基础，用表示文献学科性质的分类号来表达文献主题概念的检索语言。

分类是人们认识事物的一种思维方式，把相同与不同的事物加以区分，形成不同的类别。

“物以类聚”是人们长期以来对多种事物，包括文献信息进行识别的方法。所谓类的概念，是许多具有共同属性的事物的集合，同类的事物就称为“一类”，类的特点是具有可分性。

分类语言是建立在文献信息的可分性基础上的。

### 一、文献分类

文献分类是按照文献所载知识信息的学科属性，对文献进行类别的划分，进一步根据文献信息内容之间的内在联系组织成科学的分类体系。其目的在于使主题属性相同的文献信息集中在一起，便于检索与利用。

文献分类是在学科分类的基础上进行的，文献是人类知识的记录，学科内容是它的基本属性。学科分类，就是要依据对学科基本特征的分析，确立正确的学科分类原则，在此基础上对各门学科进行区分和排列，建立起符合现代科学发展的实际分类体系。但是，学科分类又不等同于文献分类。首先，文献分类所依据的分类法是不同的人编制的，而从事分类工作的人员在对文献进行分类时观点也有差异；其次，文献分类除了主要根据其学科内容外，还要考虑其它的特征，如：文献的类型、体裁、文种、文献涉及的时代、区域等等。此外，还要考虑不同信息单位的性质、任务以及各类型读者的需要，考虑文献内容从不同角度来利用等因素。因此说，学科分类是文献分类的基

础，它们不完全等同。

文献的分类体现在分类表上，分类表是由一系列分类号集合而成。分类号是文献信息概念的具体标识符，有文字型（如汉字、罗马字母等）、数码型（如阿拉伯数码、罗马数码等）和上述两种组合而成的混合型之分。它能反映事物的派生、隶属与平行等关系，便于检索者了解文献知识的全貌。

## 二、分类法的结构

分类法是由许多类目根据一定的原则组织起来，通过标记符号来代表各级类目和固定其先后次序的分类体系。一部完整的分类法一般由以下几个部分组成。

1. 编制说明。主要说明该分类法的编制过程，所依据的编制原则、类目设置和编制的理由，对各类分类问题的技术处理和使用，标引方法等。

2. 类目表。是分类表的主体部分。主要包括：大纲，是基本大类的一览表；简表，是一个基本类目表，它是整个类目表的构架；详表，即主表，详细列出大小类目、类号和注释；辅助表，也称复分表，是一组标准目次表，用以对主表中列举的类目进行细分。

3. 索引。是该分类法的类目名称索引，按字顺排列，以供查询，帮助不熟悉分类法的读者使用。

4. 附录。收录按类检索时经常要查阅的一些参考资料。

## 三、国内外常用分类法介绍

### （一）我国常用的分类法

#### 1. 《中国图书馆图书分类法》

《中国图书馆图书分类法》简称《中图法》，是我国图书信息界最常用、普及范围最广的一部大型文献分类法。它是北京图书馆等组织全国力量编辑而成，其编制目的是为了实现在全国文献资料统一分类编目。《中图法》由5大部类、22个大类、6个总论复分表、30多个专类复分表、4万余条类目组成了一个完善的分类体系。部类的设置采用五分的办法。为体现整部分类法的指导思想，把“马克思主义、列宁主义、毛泽东思想”作为一个基本类列于首位，“哲学”、“社会科学”、“自然科学”分别按其知识的逻辑关系列为第二、第三、第四部类，最后一个部类为“综合性图书”，收录一些内容涉及面广、类无专属的文献。

标记制度采用汉语拼音字母与阿拉伯数字相结合的混合小数层累制，以字母的顺序反映大类序列，对类目的排列采用不同的字体和行、格等形式来表示类目之间的关系。

#### 2. 《中国人民大学图书馆图书分类法》

该分类法简称《人大法》，由中国人民大学图书馆编制。《人大法》是我国新中国成立之后诞生的第一部科学性、实用性较强的图书分类法。它由4大部17大类、9个复分表和两个附表组成，用1~17的数码表示17个基本大类；对两位一类的数码标识，采用在号码数字后加小圆点，如“11.历史·革命历史”的方式，这种类号的等级排列层次分明，能从数位上看出类目间的相关性和一致性。

#### 3. 《中国科学院图书馆图书分类法》

该分类法简称《科图法》，由中国科学院图书馆编制，目前，中国科学院系统的绝大多数图书信息部门均采用此分类法。

《科图法》由5大部25大类组成，主表分两册出版，上册为《马克思

列宁主义毛泽东思想、哲学、社会科学》，下册为《自然科学、综合性图书》。其标记制度采用阿拉伯数字顺序排列制，故类目、类号的等级性不甚明确。

## （二）国外图书分类法

### 1. 《杜威十进分类法》

又称《十进制图书分类法》或杜威法，是由美国图书馆学家威尔·杜威所创制的。

杜威法是一部在国际上出现最早、流行最广、影响最大的图书分类法。它根据培根关于知识体系的分类，采用十进制的等级分类体系，把所有学科分成9大类，分别标以100~900的数字；9大类表示9个专门的主题范畴，各类中的类目均按照从一般到特殊，从总论到具体的组织原则，对不能归入其中任何一类的综合性文献资料列入第10类，即总论类，以下依次逐级分类，形成一个层层展开的等级体系。

### 2. 《国际十进分类法》

该分类法是由比利时学者鲍尔·奥特勒在《杜威十进分类法》的基础上修订补充而成，从60年代末期起被称为世界文献信息的国际交流语言。

《国际十进分类法》由主表、辅助表及辅助符号3大部分组成。它把知识分为十大门类，大类划分沿用了杜威法的基本大类结构。详表有近20万个类目，是世界上现有各种分类法中类目设置最多的一个。它的基本大类设置如下：

- 0 总表
- 1 哲学
- 2 宗教、科学
- 3 社会科学、经济、法律、行政
- 4 语言
- 6 应用数学、医学、工业、农业
- 7 艺术
- 8 语言学、文学
- 9 地理、传记、历史

该分类法类目的明细度比其它分类法高，是目前展开得最广的一部分分类法。其标记制度采用等级分明的阿拉伯数字结合多种辅助符号，列举与组配混合式结构。通过对事物整体概念的逐层分析，对特定概念进行组配，使之较好地反映多主题、复合主题的文献信息，提供了多种检索途径。

## 第四节 主题检索语言

主题检索语言的主要特征是以语词为概念标识，标识词按字顺排列，并用“参照系统”等方法辅助显示概念之间的相互关系。

用主题词作为文献信息标识有以下优点：

1. 直观性强。主题词来源于自然语言中，标识比较直观，符合人们的辨识习惯，主题词在词表中按字的字顺排列，其用法如同使用字典、词典，容易掌握，易于利用。

2. 专指性高。用作主题词的语词标识经过了全面严格的规范化处理，标识与概念严格对应，标识所表达的概念具有唯一性，故语词对概念的描述具有较强的专指性。

3. 灵活性好。主题检索语言的主要特征便是通过词与词之间的概念组配来揭示文献中的各类主题，尤其是后组式的组配原则，便于人们按照检索需要，自由组配检索概念，具有很大的灵活性。

4. 网罗性高。一个主题词表达一个事物的概念，若干个主题词合乎逻辑的组配，因此可以形成高度专指的概念特征，用于标引文献时，即可达到高度的概念网罗度。

主题检索语言的优点在于检索人员不必从知识体系的角度去判断所需文献属于什么学科，只要根据课题研究的对象，直接用能表征、描述文献内容的主题词去查检，而且同一篇文章可用多个主题词来标引，因此扩大了检索途径。

常用的主题检索语言主要有标题词语言、叙词检索语言、引文检索语言。

#### 一、标题词语言

标题词语言是一种最早出现的按主题来标引和检索文献的传统检索语言。它是以标题词作为文献内容的标识和检索依据。标题词一般分为主标题和副标题两级，标题词在编表时一一列举，而主标题和副标题已固定地组配在一起，间或在标引时再补充组配，所以标题词语言主要是一种先组式的信息检索语言。

标题法的构成原理为以下四点：

- (1) 按主题集中文献；
- (2) 用经过规范化的语词直接标引文献主题；
- (3) 用参照系统间接显示主题之间的相互关系；
- (4) 用字顺序列直接提供主题检索途径。

1. 标题词的构成。常用的标题词的类型有两种：

(1) 单级标题：一个标题仅由一个名词术语构成，可以是一个单词，也可以是一个词组。

(2) 多级标题：即复合标题，一般采用在主标题后加破折号与副标题、副副标题相结合。

一个标题词要求只代表一个概念，且必须能够直接而精确地表达文献所论及或涉及的主题，以利于提高查准率。

2. 标题词的参照系统。标题词是按字顺排列的检索系统，具有直接提供检索途径的优点，但是具有同义关系和相关关系的标题词因字面不同就不能集中在一起，从而造成同性质文献的分散，使相互间关系得不到明确的显示。语义参照系统就是为克服这一缺点而设置的。

(1) “见”参照：不用的标题词见用的标题词，它是用来揭示同义词、上位类、下位类的作用的。如“脚踏车见自行车”。脚踏车不是标题词，而自行车是正式标题词。

(2) “参见”参照：用以指引检索者从一个采用的标题词去参见与之有关的其它标题词，达到扩大检索途径的目的。这种参照可揭示标题词间的相关关系和等级关系。参见前后的标题词均是正式标题词。

(3) 注释：当有些标题词概念不十分明确或一形多义时，用圆括弧注释，对该标题词作一简要说明。

3. 标题词表的结构。标题词表是规范化的标题词词典，其体系结构一般由下列三个部分组成：

(1) 编制说明：说明该表的编制经过、收词范围、选词标准、规范化措

施、标题形式、参照系统、标引规则等。

(2) 主表：是标题词表的主体，包括全部标题词和非标题词，并有参照和注释，按字顺排列，是标引和检索时的主要依据。

(3) 辅表：是主表的辅助用表，一般用于对主表中的标题词的限定、修饰和复合主表中的标题词。辅表和主表各自分开，组成体系，在标引和检索时相互组配，构成多种标题。

## 二、叙词检索语言

叙词检索语言是应用较广的主题检索语言之一，它以规范化的名词为基础，吸收了多种检索语言的原理和方法而综合形成的一种主题法语言。叙词语言是采用单元概念的规范化语词的组配来对文献主题进行描述的后组式标引和检索语言，因此，可以说概念组配是它的主要特征。

### (一) 叙词的特性

1. 直观性。叙词使用自然语言中的语词，标识比较直观，按字顺排列，序列明确，方便检索人员了解和使用。

2. 单义性。叙词都是经过规范化处理，以达到一个叙词与一个概念严格对应的要求，故叙词概念明确，易于理解。

3. 组配性。连词组配是叙词法的主要特征。叙词较好的检索功能主要来自组配这一特性，检索人员进行检索时，只要根据检索的需要，临时从词表中选出相应的叙词，按照组配规则，任意组配检索概念，就可达到扩大或缩小检索范围的目的。

### (二) 叙词的组配

1. 组配的数学原理。组配标引是以现代符号逻辑运算为基础，而符号逻辑又是以布尔逻辑代数为基础。布尔逻辑共有三种逻辑运算：

(1) 逻辑积运算：又称逻辑乘运算，其符号为“AND”或“\*”，主要用于概念上具有交叉、限定关系的两个或多个叙词之间的运算。例如某文献系统中含有叙词“医院”的文献有：2, 4, 5, 6, 8, 17, 21 (数字表示文献号，下同)；含有叙词“信息管理”的文献有：2, 3, 5, 9, 16, 17, 20。要求检索出“医院信息管理”的文献。则可用逻辑积运算表达：

“医院”AND“信息管理”=2, 5, 17 或者

“医院”\*“信息管理”=2, 5, 17

用布尔逻辑运算图表达为图 2：



图 2 布尔逻辑图

从上图中可以看出，既含有叙词“医院”又含有叙词“信息管理”的文献共有 3 篇，即两图相交的部分，文献号码为 2, 5, 17。

逻辑积运算的结果是使概念范围缩小。

(2) 逻辑和运算：又称逻辑或运算，其符号为“OR”或“+”，是指两个或几个可能相交，也可能不相交的概念的运算。

假定 A 及 B 二个叙词要进行逻辑和运算，即表示在要检索的文献系统中，

凡含有叙词 A 或叙词 B 的文献均命中。同时含有叙词 A 及 B 的文献也命中。如上图：

“ 医院 ” =2, 4, 5, 6, 8, 17, 21

“ 信息管理 ” =2, 3, 5, 9, 16, 17, 20 则：“ 医院 ” + “ 信息管理 ” =2, 3, 4, 5, 6, 8, 9, 16, 17, 20, 21

可见，逻辑和运算是将几个叙词的文献篇数相加（重复者只计一次）。

逻辑和运算的结果是使概念范围扩大。

（3）逻辑差运算：又称逻辑非运算，其符号为：“ NOT ”或“ — ”，是指二个具有从属关系的概念运算。假定 A 及 B 两个叙词要进行逻辑差运算，则表示在含有叙词 A 的文献集合中去除含有叙词 B 的文献。

逻辑差运算的结果是使概念范围缩小。

## 2. 组配原则

（1）应恰当地从各学科中选取组配能力强、代表基本概念的科技词汇作叙词来进行组配。

（2）叙词组配是概念组配，而不是指单纯的字面组配。

## 3. 叙词的概念组配种类

（1）概念相交：两个或两个以上具有概念交叉关系的同级叙词进行组配，组配结果形成一个新的概念。

（2）概念并列：两个或两个以上具有概念并列关系的同级叙词进行组配。

叙词检索语言所具有的概念组配的特性，使得在具体的检索过程中，可随意扩检和缩检，从而体现了较大的伸缩性、灵活性和适应性，这也是叙词检索语言的主要优点之所在。

## 三、引文检索语言

引文检索语言是一种新型的信息检索语言，它是利用文献之间的相互引证关系而建立的一种自然语言，其标引词来自文献的主要著录项目。由于它与传统的信息检索语言在内容特点、检索标识、词汇来源等方面有所不同，因此引起了广大信息界及知识界的关注，并在检索实践中得到了越来越广泛的应用。

### （一）引文检索语言的产生和形成

根据引证关系编制信息检索工具源于 19 世纪 70 年代。1873 年美国出版一种称作《谢波德引文》的“法律案例索引”，其方法是以一个案例名称作为标识词，列出引用过该案例的其它一系列案例，一步步串联起大量相关的案例，从而提供众多的办理此类案例资料。20 世纪 50 年代，美国人 E. 加菲尔德对这种索引法进行了较深入的研究，并编制了一种专利索引，证明“引文法”的实用性和可行性。直到 1963 年美国《科学引文索引》编辑出版，报道了 1961 年内出版的 613 种重要科学期刊发表的文献 113318 篇，及其引证的 1370000 篇文献间的相互引证关系，从而获得巨大成功，也使得《科学引文索引》成为世界上最主要的检索工具之一。

引文检索语言正是利用文献之间的“引证”与“被引证”的关系建立起来的。文献大范围内以“引证”与“被引证”关系串联起具有一定相关程序的“著者网络”和“文献网络”，以此原理出发，进而扩大并研究其中的关系，并对其间的规律性加以阐述和证明，用于文献信息检索工作，即形成独具特色的新型信息检索语言——引文检索语言。

## （二）引文检索语言的特点

1. 其主要检索标识为被引文献的著者姓名。著者姓名不仅是文献的外形特征，说明某文献由某人撰写，而且从一定意义上说也是文献的内容特性，即著者姓名也能够说明文献的学科属性。如某人因在某方面的成就而享誉该领域，或者因其特殊的贡献，以其发明、创造等命名，故出现了一些行业中的人名代学科、专业的现象，因此，著者姓名作为一种检索标识在西方图书信息界已较普遍。

2. 选词方便。引文检索语言的检索标识词来自文献本身，无需规范，也无需词表。文献撰成时著者即署出姓名，加以标识，所以说文献的标识由著者提供，与文献同时出现在加工标引人员面前，而且著者姓名一般不大变化，具有较好的稳定性。由于它来自自然语言，可随时从文献中选取。

3. 词汇丰富。引文检索语言源于“追溯法”，但优于“追溯法”。它不是以单一文献为起点，而是以一定范围内出版的众多新文献一同多向追溯，因而联系的文献多，著者多，标识词也多，检索时可随手从加工的文献中选取。它检索面广，易于从较大的范围内发现同性质的文献信息。

## （三）引文索引的组成部分

引文索引由三个来源相同的部分组成，但它们具有不同的排列方式。

1. 引文索引。列出一段时间发表文献的全部被引文献，按被引文献的第一著者排列，其下按时间先后列出各被引文献的出处，包括年份、刊名、卷次、页次等，再在各条被引文献之下依次列出引证过它的全部文献。引证款目以第一著者姓名的字顺排列，再列引证文献的出处，包括刊名、卷次、页次、年份等。在引证文献和被引文献的出处部分均不列出篇名。

2. 来源索引。是引文索引最基本的部分，因为从另两部分初步查到的线索都要到此部分从篇名上加以核实。这一特点使其类似传统检索工具中正文部分的作用。来源索引按引证著者姓名的字顺排，每个款目可包括姓名、文种代号、篇名、刊名、卷次、期次、页号和发表年份。此外，还列出参考文献的数量和第一著者的联系地址。

3. 轮排主题索引。此部分是以文献篇名的关键词轮排主题索引，相当于一般关键词索引，只是在配词时予以限定，以避免过于繁琐。其控制的方法是限定一些词只能作配词而不能作为实词进行搭配，这就相对地减少了一些词的轮排机会，保证索引体积不至臃肿，检出不至过泛，从而保证检索效率的提高。



## 第四章 信息检索工具

信息检索工具又称文献信息检索工具，其目的是为科学研究活动提供文献检索服务。因为它提供的不是有关数据、事实或资料性的信息，而是文献信息的线索和指引，故又被称之为线索型检索工具。

### 第一节 信息检索工具的特点

检索工具是指用来存贮、报道和检索文献信息的工具，它是为了检索需要而组织的文献信息检索性系统。它一方面是索引工作经过搜集、加工、标引、组织文献的结果；另一方面又是寻检工具赖以进行的主要工作手段和条件。

文献信息检索工具以各种原始文献为素材，在广泛收集并进行筛选后，分析和揭示其外形特征和内容特性，给以书目性的描述和来源线索的指引，形成一定数量的文献信息单元，再根据一定的框架和顺序加以排列或形成可供查检的卡片或工具，或以图书的形式出版，或以期刊的形式连续出版，成为向社会进行广泛报道的二次文献，使科研人员从中了解本专业学科或领域的进展情况及科学技术发展的全貌。同时，还可以了解图书、期刊等各类文献的出版情况及其在一些图书信息部门的收藏情况，易于利用。

概括起来，文献信息检索工具主要具有以下特点：

1. 详细而又完整地记录所著录文献的线索，包括文献篇名、著者、来源出处等项目，以使用户能利用这些线索找到有关文献。也就是说，文献信息检索工具的收选内容是文献信息单元，不是经过浓缩加工后的知识单元，更不是原始文献，它给予用户的是有关文献的信息线索，如用户欲获得所需的知识，则还必须通过这些线索到所指引的一次文献中去浏览并加以选择；同时，这些文献信息单元构成了检索工具的主体内容。

2. 所提供的文献信息的线索准确可靠，内容丰富系统，其所收录的文献信息来自于原始文献，且选题比较灵活实用，以用户需要为选材原则。

3. 对所著录的文献，都标有可供检索的各种标识，如分类号、主题词、文献序号、代号代码等等，以使用户利用这些标识来检索所需的文献。

4. 在体例编排结构上，从实用易检出发，可以结合文字特点和学科特点对所选的款目按分类排组或按主题、叙词、关键词等的字序排组，并利用“参照”关联相关各个部分。此外，又辅以适宜的辅助工具，以便同主体的排列相辅相成。

5. 提供丰富的检索手段，即各种体系的索引，如分类索引、主题索引、作者索引、代码索引等，以使用户能从各种途径方便地进行检索。

6. 出版形式多样性，可以是图书、期刊、卡片、缩微品、磁带、磁盘、光盘等，兼备对文献信息的揭示报道、存贮累积和检索利用的功能。

### 第二节 信息检索工具的类型

文献信息检索工具种类繁多，数量庞大，且表现形式各异，因此，可以从不同的角度对文献信息检索工具进行分类。

#### 一、按检索手段或设备划分

### （一）手工检索工具

指以普通卡片目录和书本式文摘、索引等形式，以手工方法检索为基础的文献信息检索工具。其历史悠久，至今在我国仍是最主要的检索方式。手工检索，几乎不需要任何设备，查找方法简便灵活，检索费用低，但费时费力，检索效率低，检索效果差。

### （二）机械检索工具

指运用一定的机器设备来辅助检索文献信息的检索工具。主要有机器穿孔卡片检索工具和缩微文献检索工具。

以穿孔卡片为载体的检索工具，是手工检索到机械检索的过渡。最早的手检穿孔卡片检索工具出现于1904年，后来发展到边缘穿孔卡片、比孔卡片到机械穿孔卡片等。但是自计算机检索出现后，穿孔卡片检索工具已逐渐不再单独使用。

缩微文献检索工具又称光电检索工具，它是文献缩微品作为文献库，用一定的光电设备从中进行文献信息检索。一张缩微平片可以缩摄存贮几十页至几千页的文献，且存贮时间较长，已普遍运用于一些珍贵文献的复制保存。

### （三）电子计算机检索系统

它是磁性介质为载体，以计算机来处理 and 查找文献的一种电子化自动化系统，由计算机、检索软件、文献数据库、检索终端及其它外用设备组成。用户可以通过终端设备和通讯线路与相关检索系统联系，查找所需文献。电子计算机检索的速度和效果都明显优于其它检索方式，目前在世界各国都已得到了迅速发展。

## 二、按物质载体形式和种类划分

### （一）书本式检索工具

书本式检索工具又可细分为期刊式、单卷式和附录式三种。

1. 期刊式检索工具。它是在一个名称之下，定期连续刊行的一种检索工具，具有期刊的特点，是目前查找科技文献的主要检索工具。其优越性在于收录文献新，报道文献快，且能够比较完整系统地收选一个学科领域的有关文献信息，便于回溯检索和全面了解该学科领域的发展状况；同时也便于装订、保存、借阅、管理。

2. 单卷式检索工具。这种检索工具大多是以一定的专题为内容而编印、单独出版的。它收集的文献比较集中，往往积累了一段相当长时间的文献，并以特定范围的读者作为对象。对于专题文献检索比较方便，有较高的使用价值。

3. 附录式检索工具。这种检索工具不独立出版，而是附于有关书刊之后，但具有一定的参考价值。尤其是作为情报信息研究成果的综述、述评所附的参考文献目录，往往是通过全面搜集大量文献进行精选而成，所以具有较大的价值，也越来越受到人们的重视。

### （二）卡片式检索工具

它是文献收藏单位揭示馆藏文献信息的常用检索工具，如图书馆目录，它把每条款目写在或印在一张卡片上，然后按一定的方式将卡片一张张排列起来，成为成套的卡片。一般包含主题目录、分类目录、篇名目录、著者目录等。其优点是可以随时抽排，不断充实、更新，及时灵活地反映现有文献信息。

其缺点是占有较大的馆藏空间，体积庞大，成本费用也比较昂贵，制作费时费力等。

### （三）缩微型检索工具

它是指计算机输出的缩微品，有平片与胶卷两种形式。由计算机将存贮在计算机存贮器里的书目著录，按照人们指定的格式和排列系统进行输出，一张普遍的缩微平片可包含 3000 多条书目著录，即能代替 3000 多张卡片。

其特点是存贮量大，体积小，成本低廉，易于保存。但它不像卡片式检索工具那样可随时增减款目，需由计算机进行全套更新，所需费用较大。

### （四）机读式检索工具

这是将书目著录按照一定的代码和一定的格式记录在特定载体上，专供计算机“阅读”的检索工具。只有借助于计算机，才能对它进行检索。例如，一盘规格为 2400 英尺的机读磁带，可记录 4000 万个字符，相当于每页 6000 字的文献 6600 页，而记录时间仅需 20 ~ 30 分钟，并可实现多种形式的输出，如在计算机上显示出来，或用打印机打印，还可以存贮在个人磁盘中保存等。其优点是查找文献迅速准确，检索效果好。

## 三、按收录的内容范围划分

### （一）综合性检索工具

### （二）专科性检索工具

### （三）专题性检索工具

### （四）单一性检索工具

## 四、按文献的著录方式划分

### （一）目录

### （二）题录

### （三）文摘

### （四）索引

### （五）综述

## 第三节 信息检索工具的结构

文献信息检索工具收录的不是原始文献本身，而是有关原始文献的信息和指引性线索，因此，检索工具对文献的揭示有其自己的特点，对文献的著录和标引也有一定的方式，而且在体系结构上形成了比较固定的编排形式。

### 一、信息检索工具的文献著录及标引

文献的著录是对文献的外形特征进行确切的描述；文献的标引是对文献的内容特性加以准确的标识。

#### （一）检索工具的文献著录方式

1. 图书的著录。一般分为两种情况处理，一是如为整书著录则取书名、著（译、编）者及各出版事项等；二是如为书中选出的章节著录，则以章节名作为篇目名，再附以图书的出版事项以及该章节的起止页号等项加以著录。

2. 期刊文献的著录。著录内容包括：论文篇名、著者、著者工作单位、刊名、年、卷、期、页号等，如属外文者则加译篇名，并将原文篇名加括弧附于其后，并注明国别或文种。

3. 会议论文的著录。可仿照图书章节的分析项目，即：论文篇名、著者、

会议录名称、会议日期、地点、起止页号、文种以及出版事项等；如为整本的会议录，则仿图书处理，附著有关会议的事项。

4. 科技报告的著录。一般著录报告的篇名、著者、工作单位、报告号、年、月、日及文种。如系外文，则加译篇名，并在其后注出原文篇名。

5. 学位论文的著录。要求著录论文题目、著者、导师姓名、大学系科名称或机构名称、专业名称、学位级别、颁发学位日期、出版日期、页数、学位编号、论文存贮符号等。

6. 专利文献的著录。包括专利说明书题目、发明人、专利国别、专利号、申请日期、公布日期等。

另外，还有政府出版物、标准文献、档案、产品样本等文献的著录，可基本仿照以上文献的著录方式。

## （二）检索工具的文献标引方式

文献的标引是对其内容的学科特性加以反映标识。一般处理的办法是采用相关的分类表或词表作为依据，将文献的学科属性或其实际内容加以分析鉴别，并作出标识，以便存贮于一定的检索系统，以待查检。这类标识可以是类目、类号、主题词、叙词、单元词等，也可以是根据题目、文献的内容摘要或文献全文中选出的能够代表其内容性质的重要词汇。这类的类表、词表或词汇，连同其使用规范和注释等统称为信息检索语言，或称为文献语言，本书上个章节已有介绍。

### 二、信息检索工具的结构

文献信息检索工具的结构同信息检索的效率和效果有很大关系。在长期的工作实践中，其体系结构已趋于定型，一般由主体、辅助工具、使用说明和附录四个部分组成。

#### （一）主体部分

主体部分是检索工具的主要部分和基础部分，由所收录文献的著录款目组成，是存贮与检索文献的实体。其具体编排是将经过著录和标识的诸多文献的条目依据所加的标识，把性质相同者集中，再按一定的体系原则组织起来，而对于内容相近或相关的款目或标目未能置于相关位置者，则以“参照”项目指引到所采用的标识款目处或标目处，以使主体部分成为一个具有内在联系的有机整体。

#### （二）辅助工具

主要是指主体部分的辅助索引。它是在文献检索工具中，从主体部分的文献款目编排系统以外的角度，增加检索途径，扩大检索效果的辅助性手段。如一些检索工具中附带的主题索引、著者索引、文献类型索引、分子式索引、生物系统索引等。它不能离开主体部分而独立存在。

#### （三）使用说明

使用说明是检索工具的重要组成部分，主要包括前言和后记，用以说明检索工具的编制目的、用途、收选文献的学科范围、时间范围、地域范围、著录项目和格式、编排方法等。除文字性的说明外，一般从书中选出典型例子加以解释和示范，以增加直观效果，便于理解和掌握，目的在于帮助用户正确利用检索工具，提高检索效率。

#### （四）附录

附在检索工具后面的有关参考资料被称为附录。它多附于全书之后，数目多少不等，但都是与书的内容有关，可供经常查检使用，如书中采用的符

号、缩略语、计量单位等。

#### 第四节 检索工具的主要形式

##### 一、目录

###### (一) 目录的概念

目录是经过著录的一批相关文献，按照一定的次序原则编排而成的一种揭示与报道文献信息线索的工具。目录多以图书的形式单卷出版或多卷连续出版，提供广大读者使用参考。

目录的特点是以文献的自然出版形式，如整本书、整张图、整份资料为著录单位存贮和供检索的。但由于目录只是描述出版物的基本特征，所以对出版物内容揭示程度比较浅，例如图书馆目录，一般主要用来检索图书馆有没有和有哪些自己所需要的图书、期刊等。

目录起源于早期图书馆用以揭示报道其藏书情况的卡片式目录，后经逐渐发展演变成为当前常见的具有广泛内容和多种出版形式的书本式目录。现代科学技术的发展推动了目录工作的实施，并不断演化形成了独立的检索工具系列，在文献信息的检索中发挥着系统性回溯和报道新文献的重要作用。随着计算机技术的广泛应用，文献工作标准化和协作化的发展，“缩微型”目录和“机读型”目录的出现使目录工作发生了重大的变化，从而也推动了目录工作的进一步发展。

###### (二) 目录的类型

目录具有不同的类型，可以按照不同的标准加以区分。

1. 按其收录文献的内容范围，可分为综合目录、专题目录、地方文献目录、个人著述目录等。
2. 按其收录文献的类型，可分为图书目录、期刊目录、资料目录等。
3. 按其编制目的和社会职能，可分为登记目录、科学通报目录、推荐目录、专题目录、目录之目录等。
4. 按其反映文献收藏状况，可分为总目录、馆藏目录、联合目录等。
5. 按其物质形态，可分为卡片式目录、书本式目录、机读型目录等。
6. 按出版物的语种，可分为中文目录、西文目录、日文目录、俄文目录等。

目录的类型还可以根据其它的方式加以划分，但在实践工作中常用的目录工具主要有图书目录、期刊目录、专题目录、目录之目录等，下面分别对这几种目录加以简介。

(1) 图书目录：以图书为收选对象，又称为书目。此类目录因同图书馆的藏书和出版业的出版目录关系密切，又可细分为馆藏目录、营业目录、推荐目录和联合目录等，还可按收选的学科范围分为综合性书目、专题性书目、专科性书目等。由于有些目录工程浩大，又多由政府指定的国家图书馆或出版管理部门统一进行，这种系统而又完备的目录称之为国家书目。除回溯历史上出版的图书外，对新书采用登记的办法加以法定的控制和掌握，所以当前对于近期的国家书目也称作登记性书目。另外，各图书馆和出版部门又出版以报道新书为特点的报道性书目或推荐性书目。

(2) 报刊目录：以某个图书文献部门收藏或某些出版部门出版的期刊、报纸为主要收选对象，形成综合性、专科性或专题性的目录。它也可以形成

某个国家、地区或某种文字的报刊目录，或者具有收藏性的馆藏报刊目录、专科报刊目录，以向读者进行详细报道和提供检索使用。

(3) 专科或专题文献目录：以期刊、图书、会议录中的某一专科或某一专题内容为对象，适当增加专利文献、科技报告及单位论文等，也可加上整本的有关图书或会议录等，形成某一专科领域的文献目录，或某一专题性质的文献目录，以供读者集中了解本专业范围的最近文献，满足特定领域的文献信息需求，具有较大的参考价值。

(4) 目录之目录：又称为书目之书目或目录指南。它是多种检索工具汇集在一起而编成的一种特殊的目录类型，以提供掌握各种目录的工具。读者可以通过它了解到目前已出版的各种目录工具的性质和特点，以供选择使用。国外很重视编制目录之目录，如美国的《世界书目之书目》、日本的《日本书目之书目》等，都比较著名。

### (三) 目录的著录及编排

目录的著录款目通常包括以下各项：

- (1) 书名项或刊名项；
- (2) 著者项；
- (3) 出版项：包括出版者、出版地点、出版日期、版次等；
- (4) 稽核项：包括页数、图表、开本、装订形式、定价等；
- (5) 附注项：对上述各项的附注或说明；
- (6) 提要项、内容简介；
- (7) 业务注记：指图书馆在目录上做的业务记载，包括索书号、登录号、分类号、主题词、存贮地点等。

目录的著录在于揭示所收录的文献的基本特征，以利于方便检索。而不同的目录工具有不同的编排形式，能提供不同的检索途径。

在目录体系中，书名目录、著者目录、分类目录和主题目录是目录检索工具提供的最常用检索途径。下面对其各自的排检方法加以简介。

#### 1. 书名目录

(1) 中文图书：按书名字顺排检，具体方法有汉语拼音法、笔划法、部首法等。

(2) 外文图书：西文书名按字顺排检；日文书名有日语汉字和假名两种，前者按汉字笔划排检，后者按假名字顺排检；俄文书名按字母顺序排检。

#### 2. 著者目录

(1) 中国著者：按姓名字顺排检。

(2) 外国著者：按姓在前名在后排检。而欧美人姓名排列习惯是名在前姓在后，故编目时应将著者的名姓倒置，中间用逗号分开。

3. 分类目录。分类目录是按分类法规定的图书、资料所属的学科、专业内容对应的分类号编目，并按类号的顺序排列。

4. 主题目录。主题目录是图书、资料的主题内容编排的目录，按主题词字顺排列。

## 二、索引

### (一) 索引的概念

索引是将图书、期刊等文献中的一些重要的、有检索价值的知识单元，如主题词、分类号、著者姓名、名词、公式、数据、事实、机构、论文篇名等，根据需要一一分析摘录出来，并注明它们所在的页码和文献号，再按一

定的顺序编排组织起来，构成检索的种种途径，这种检索工具称为索引。

索引与目录相比较，在本质上是一样的，即是一种传递文献信息、揭示和检索文献的工具。但索引在结构和功能上有自己的特点，主要表现为：

1. 侧重揭示文献的内部特征。目录重在文献的宏观描述，对其外部特征给予详细而精确的揭示。而索引则侧重文献的微观描述，对其内部的特定事项或单元知识给予具体而深入的揭示，可从内容深度方面描述读者的要求。这种检索深度是目录所不能达到的。

2. 著录内容标明出处。索引的款目有指引项，以指示具体标目所含文献信息在原文中的确切位置，这是索引区别于目录及其他检索工具的重要特征。

3. 揭示范围广，信息含量大。目录的著录对象仅限于图书及其外部特征，而索引的著录范围涉及到书、刊、报等各类文献的各种信息，容纳了较丰富的信息量。

## （二）索引的类型

索引的出版数量大，种类繁多，且各具特色和功能。索引大体可以分为几大类型：一是按文献外部特征编制的索引，有书、刊、篇名索引，著者索引，文献序号索引等；二是按文献内容特征编制的索引，有分类索引、主题索引、关键词索引等；三是按索引所反映和揭示的文献类型划分，有图书索引、期刊索引、专利索引等；四是按标引语言来划分，有主题索引、分类索引、著者索引、号码索引等。下面着重介绍几种常用的索引。

1. 主题索引。主题索引是将文献中具有实质意义的词语或能揭示文献主题概念的词语抽出来，经过规范化处理之后，再按字顺排列起来组织标识系统，或在各主题词下面给出副标题词、文摘、文献出处，或在各主题下面给出篇名性的说明语，或关键词性的说明语，然后在说明语的后面列出文摘号，以这种方式编制的索引称为主题索引。

主题索引给检索者提供从研究课题的主题概念出发查找文献线索的途径。主题索引又可分为以下几种：

（1）标题词索引：标题词是自然语言中比较定型的事物名称，是经过规范化、用以表达文献主题内容的词、词组或短语。标题词索引就是以标题词组织成的索引系统，它是最早的一种主题索引。通常，利用标题词索引，要借助于该检索工具规定的标题词表，以便选准标题词进行检索，从而取得好的检索效果。

（2）叙词索引：叙词又称主题词，是从文献中优选出来并经严格规范化处理的词。它能表达文献内容特征，词间有严密的语义关系，又具组配性能，叙词索引即是按叙词字顺排列而成的。它也是一种概念组配索引，通过叙词间的概念组配表达文献主题和信息提问。使用叙词索引时要利用叙词表，以正确选择检索词。叙词索引特别适用于电子计算机检索。

（3）关键词索引：它是由文献中具有实质意义的词（即关键词）按字顺排列而成的。一篇文献被抽出的若干关键词，每个都轮流排在首位，因而从各个关键词着手，都能查到该篇文献。但由于关键词系自然语言，又未经规范化，各词之间也无语法关系，因此容易出现漏检、误检现象。

（4）单元词索引：单元词是经过规范化的、能表达文献主题的最小概念单元的词，是不能再加分解的独立词汇。它是一种概念组配索引，即通过最基本的词汇（单元词）的组配来表达复杂的概念。检索时，根据提问，把有

关的概念单元组配起来进行查找。但由于单元词本身存在的缺点，如强调词的单元化，仅是词的字面上的分解和组配，有的词无法单元化，无参照系统来反映单元词间的关系等，因而大型索引及检索系统一般不采用单元词，而采用功能更强的叙词，所以单元词索引已经较少应用了。

2. 著者索引。著者索引是以文献署名的姓名（注：欧美人按名在前姓在后）的字顺排列，并在著者名字后面列出文摘号，以这种方式编制的索引称为著者索引。

著者包括个人著者、团体著者、专利发明人、专利权所有人等，因此著者索引也就包括个人著者索引、团体著者索引、专利发明人和专利权所有人索引等。著者索引为检索者提供从已知著者姓名入手查找文献线索的途径和方法。

3. 分类索引。分类索引是将文献内容所属的类号按隶属关系排列起来，并列与分类号相应的类名，以这种方式编制的索引就称为分类索引。它是科学体系分类为基础，按照一定的观点和原则，结合文献的特点，采用概念划分方法，将文献组成具有展开、隶属关系的索引体系。

分类索引给检索者提供从学科分类角度查找文献线索的途径。利用分类索引时，需先熟悉和了解该索引的分类方法，才能取得较好的检索效果。

### 三、文摘

#### （一）文摘的概念

文摘是以简明扼要的文字摘述文献的主要内容和原始数据，向用户报道最新研究成果，传递文献信息和查询文献线索的一种工具。它著录文献外部特征，更着重描述文献的内容，是原文浓缩后的产物，是系统报道、积累、检索一次文献的主要检索工具，是检索工具的主体，二次文献的核心。

文摘是随着文献的急剧增加而产生的。由于科学技术的发展，文献数量大增，读者在有限的时间内难以逐一阅读原文，而传统的检索工具又不能及时地报道、揭示最新的文献内容，因此，文摘应运而生，并越来越得到人们的重视。

与其它检索工具相比，文摘主要具有如下特点：

1. 浓缩性。文摘是对原文的高度浓缩，它以凝炼的文字表述完整的信息，以最小的篇幅提供最大的信息量，使文摘在一定程度上起到原始文献的作用。

2. 客观性。文摘是对原始文献的简明的真实的复述，它忠实于原文的本来面貌，客观准确地摘录它的内容要点，不做说明解释，更不进行引申评论。

3. 检索性。文摘不仅有规范化的编制系统，标引和著录格式，还具有引导性和检索性。文摘注明文献的出处，提供检索原文的线索，并且提供检索用的辅助索引，如主题索引、著者索引、年度累积索引，使读者从不同的角度检索刊物中的每一篇文摘。

#### （二）文摘的作用

概括起来，文摘具有以下几个方面的作用：

1. 报道最新科学文献，使读者及时了解本专业最新文献出版情况和研究进展。

由于文摘所摘录的是经过筛选的某一学科或某一专业领域最新最有学术价值的文献，并把相关论题集中一处，因而通过一组文摘，能获悉该学科或专业领域的学术概况及最新进展。特别是首次文摘，能使最新科研成果及时



公之于世，时效性强。

2.可以帮助读者判断检索的文献是否合乎需要。文摘能够简明扼要地揭示文献的主题内容，因此，读者在不需阅读原文的情况下，就可真实地了解该文献的内容，明确其基本要点，从而较容易地判断该文献是否为自己所需。

3.可以帮助读者消除语言上的障碍。一般读者不可能掌握多种语言，而文摘性检索工具收录的文献都译成一种常用的文种，即利用这一常用的文种就可以阅读到大量的文献，从而在一定程度上消除读者的语言障碍。

4.可以帮助读者对一些无法找到的文献有大致的了解。阅读了文摘大体上就掌握了文献的主要内容，这就使读者有机会大致了解自己难以获得的相关文献。

5.可以节省阅读时间。由于文摘是对原文献主要内容的描述，每一条文摘实际是一篇高度浓缩、信息完整的文献，所以读者阅读文摘的时间则远远要少于阅读原始文献所花费的时间。

6.引导读者查找原文。文摘具有引导性及检索性，读者可利用文摘对原文献出处的注明，掌握其来源和线索，并据此查阅对自己具有参考价值的原始文献。

### （三）文摘的重要类型

文摘的类型有许多种，但最常用的还是按照文摘编写方式及其对文献的揭示程度来划分的三种主要类型：指示性文摘、报道性文摘和简介性文摘。

1.指示性文摘。指示性文摘是把原文的主题范围、目的、方法指示给读者，不直接摘录原文的论点和数据。它主要适用于那些篇幅过长、内容较散的文献，如泛论性或评论性的文献等，一般不涉及具体的技术问题。其字数一般在100字左右，简短扼要，起到题解作用。

2.报道性文摘。这类文摘是在对原文献进行深入的语义和逻辑分析的基础上进行高度浓缩而形成的。它概括地叙述原文献所有或部分重要信息，包括研究对象和目的、观点和方法、主要结论、全部论据及其价值等。报道性文摘所含信息量大、参考利用价值高，在一定程度上能够取代原始文献。该类文摘的篇幅一般在300~500字之间。

3.简介性文摘。该类文摘仅仅著录文献的外表特征，涉及文献的内容性质浅，或仅提供简单的说明性和注释性的文字。

### 四、综述及述评

综述、述评属于三次文献范畴，是检索工具和非检索工具之间的一种文献，具有一定的存贮、报道和检索功能。

综述是综合、评述的简称，以述为主，述中有评，是作者研究了某一课题大量有关文献后对该课题在一段时间内发展情况所做的精练的综合叙述与分析。它是几十或几百篇文献信息的高度浓缩，能够比较全面而又深刻地反映某一专题的发展、水平、动向，具有指导意义和较高的参考价值。

述评是以评为主，评述相兼，要求更高的一种文体。它以较小的篇幅就某课题现状、进展进行描述、评论，并提出方向性或指导性意见，具有较高的权威性。

综述和述评都是由本学科领域的专家或权威人士撰写，是具有重要价值的文献。它所附的大量参考文献又是查找该学科领域重要文献的线索，可以使读者对该学科领域作进一步的了解和深入研究。

综述、述评一般发表于该学科、专业领域的核心期刊中，也有以书、刊形式加以集中出版，以供读者系统查阅。

## 第五章 中外文检索工具举要

### 第一节 中文文献检索工具概况

科学技术的发展和文献的迅速增长推动了文献检索工作的进步，也促进了文献信息检索工具的加速出版和检索体系的建立，从而为广大科研工作者快速有效地查找文献信息提供了便捷的途径。

我国的文献检索体系的建设基本上是以建国后开始的，并经历了四个不同的发展阶段，逐步形成了比较完善的检索刊物体系。

#### 一、我国文献检索体系建设的发展阶段

1. 1949~1961年，是从无到有的发展阶段。这个阶段经历了两个“从无到有”的转变。

第一个“从无到有”是从没有文献到开始组织力量翻译出版前苏联文摘杂志。1956年，中国科技情报所成立，当年便组织人员翻译出版了前苏联的文摘杂志中的两个分册，即《机械制造文摘》和《冶金文摘》。到1961年，翻译出版的前苏联文摘增加到50种（分册），约占前苏联出版杂志的二分之一。同时，铁道科学研究院出版了《铁道文摘》，主要翻译的是英、美、德、日、荷等国铁道运输方面的文摘。

第二个“从无到有”是1958年我国开始出版题录性的索引刊物《期刊论文索引》（后改名为《科技文献索引》），主要报道国外期刊论文，从而实现了从没有索引到有少量的主题索引。此期主要是翻译出版国外的检索刊物，报道速度十分缓慢，且选题也不尽切合国情需要。因此，在1961年，国家科委成立了“中国国外科学技术文献编译委员会”，负责组织协调全国的科技文献编译报道工作，并决定“由全部翻译苏联文摘逐步过渡到自编世界各国的文摘”，进一步促进了我国检索刊物的发展。

2. 1962~1966年，是从翻译为主向“自编本”过渡，我国检索刊物体系已具雏形的稳步发展阶段。

此期检索刊物的出版编辑工作基本上是有组织、有计划的，编辑单位主要有中国科技情报所及其重庆分所和各地方专职情报所。这一阶段，我国先后出版了《中国机械文摘》、《中国化学化工文摘》，以及报道国外文献的《科技文献索引》等刊物。到1965年，已出版自编的检索刊物30个分册。这些检索刊物，除报道我国的科技文献外，也报道美、英、法、德、日、苏等国的科技文献，每期报道量为几百至一千多条左右。到1966年6月，全国出版的检索刊物已达59种，112个分册，其中题录性的有29种，29个分册，年报道量40多万条；文摘性的有30种，83个分册，年报道量35万条。科学技术文献出版社计划出版27种文摘的年度主题索引。除此之外，还出版有专题性、单卷本的文摘、题录等检索工具。

可见，这个时期我国的检索刊物体系已初具规模。

3. 1967~1976年，是从有到无的倒退阶段，我国文献检索刊物的编辑工作基本停顿。

这个阶段，我国绝大多数文献检索刊物相继停刊。其间，虽有少数检索刊物复刊，但缺乏组织和统一规划，有的出出停停，有的则复而又停，造成检索刊物质量较低，缺乏价值。

4. 1976年至今，是高速发展阶段。1977年7月，在石家庄召开了“全国

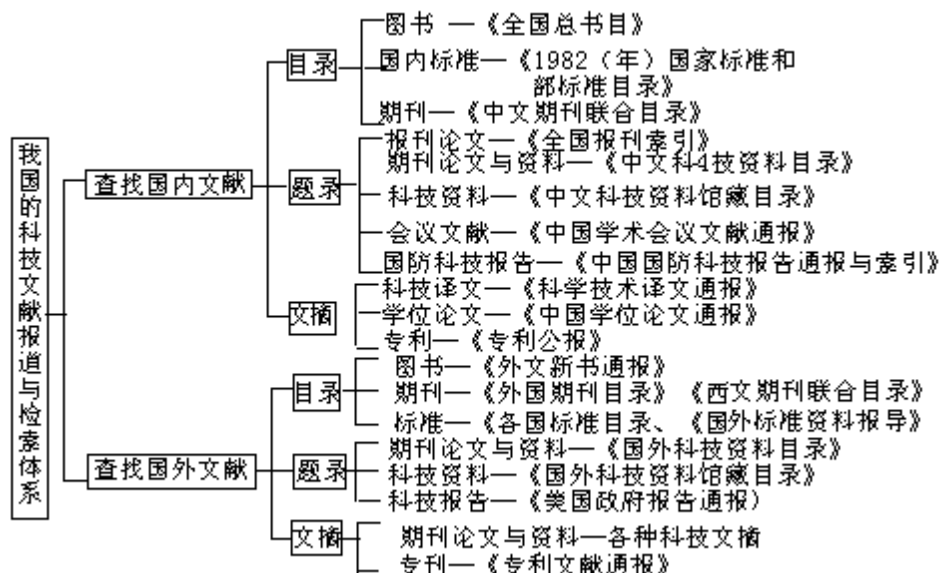
科技情报检索刊物协作会议”，议订了“关于建立健全我国科技文献情报检索刊物体系的意见”和“1979~1985年全国科技文献编译出版规划”(草案)，开始了有计划、有组织、有领导地统一建立我国检索刊物体系。首先有组织地创办了报道国内文献资料的《中文科技资料目录》，共22个分册。1979年，中国科技情报学会编辑出版委员会，在山东淄博市召开了学术会议，会上决定成立“中国科技文献检索刊物总编辑部”，下设10个分编辑部，统一规划、统一选题、统一编辑、统一标引，统一出版发行各种检索刊物。到1985年，我国公开发行的检索刊物已达106种。

目前，我国已经基本形成了一个较完整的检索刊物体系，并逐步走向计划化、标准化和协调化，具备如下特点：出版形式和内容有了统一规划；

学科覆盖面越来越广，专业范围越分越细；标引工作和著录格式日趋标准化；检索手段和检索途径逐渐完善；一批专门报道国内文献的文摘刊物出现；报道量持续增长；编辑单位由专职信息机构部分转向专业信息机构，提高了检索刊物的内在质量；电子计算机技术广泛应用。

## 二、我国文献检索工具体系

可用图3展示我国的文献报道与检索体系：



## 第二节 中文图书检索工具

图书是最常见的文献类型之一。检索中文图书，主要利用一些综合性书目和各图书馆的藏书目录。常用的图书目录主要有以下几种。

### 一、《全国新书目》

《全国新书目》由中国版本图书馆编制、中华书局出版的查找国内出版的图书的总目录。它是根据全国出版单位向版本图书馆缴呈的出版物样本编成的目录，它比较系统地反映了我国图书出版的全貌，是检索国内图书的主要检索工具。

《全国新书目》以月刊的形式出版，按月报道我国已出版的最新图书。它的编排方法采用分类编排，类目包括：马克思列宁主义、毛泽东思想；哲学；社会科学总论；政法；军事；经济；文化；科学；教育；语言；文学；

艺术；史地；自然科学；工业技术；环境科学等 22 个类目。它所分类的类目较粗，查找时需在大类下逐条查找。其著录的内容包括大类名称、小类名称、书名、著者、出版社、出版年月、页码、开本、定价等。

## 二、《全国总书目》

《全国总书目》是《全国新书目》的累积本，由国家版本图书馆编辑出版，全面系统地报道全国当年出版的各种图书。和《全国新书目》相比，它的功能在于累积检索，而《全国新书目》的功能在于及时报道出版的新书。二者可以互相取长补短，选择使用。

《全国总书目》由分类目录、专题目录和附录三部分组成。

分类目录是《全国总书目》的主体部分，其类目设置同《全国新书目》一样，共分 22 个大类，大类之下再分成小类。其收录对象全部是中文出版的图书。

专题目录包括技术标准目录；少年儿童读物目录；中小学、师范学校、业余学校教学用书目录；少数民族文字图书目录等。

附录部分包括报纸、杂志目录；出版者一览表；书名索引三个部分。

检索全国出版的中文图书时，可以通过分类目录查找，也可以在知道书名的情况下，利用附录中的书名索引查找图书。可以说，《全国总书目》是查找全国出版的中文图书最全面、最主要的检索工具。

## 三、《科技新书目》

《科技新书目》由新华书店北京、上海发行所主办、北京发行所编辑出版，主要报道京、沪、津、渝等地出版的科技图书信息。它编入的科技图书占国内出版的科技图书的 70% 左右，包括新书预告、征订书介绍、存书介绍、重要图书出版消息等栏目。

《科技新书目》是为了通报即将出版的图书信息，提供图书订购的依据。虽然它属于非检索类刊物，但通过它可以了解国内科技图书的出版动态，为读者提供检索的线索。

## 四、《外文新书通报》

《外文新书通报》是检索外文图书的检索工具，报道馆藏的外文新书。有中国科学院图书馆、北京图书馆、安徽省图书馆、江苏省图书馆等单位编辑的不同版本。中国科学院图书馆为报道最新馆藏的外文图书，编辑出版了《外文新书通报》，1973 年创刊，主要报道西文、俄文、日文的社会科学、自然科学和工程技术方面的图书。

另外，外文图书的检索工具还有《上海市外文新书联合目录：自然、技术、科学部分》，由上海图书馆编辑出版，报道上海市有关图书情报单位所收藏的外文原版图书，包括西文、俄文、日文等文种。

### 第三节 中文期刊资料检索工具

期刊资料是极其重要的信息源。据估计，来自科技期刊的信息占整个信息来源的 65%。目前，世界上出版的科技期刊有 45000 多种，每年发表的期刊论文达 300 万篇以上，而我国出版的中文期刊杂志在 1988 年就达 5865 种，其中科技期刊 2877 种，因此，科技期刊的检索工具已成为科研人员进行科学研究的重要工具。下面对常用的期刊资料检索工具加以简介。

#### 一、《全国报刊索引》

《全国报刊索引》由上海图书馆编印；1951年创刊，月刊。其前身为山东省图书馆编印的《全国主要期刊重要资料索引》。从1955年起由上海报刊图书馆（现为上海图书馆）接办，1956年改名为《全国主要期刊资料索引》，在1966年以后停刊，1973年10月复刊，改为现名。

1980年，《全国报刊索引》分为自然科学技术版（科技版）和哲学社会科学版（哲社版）。它是根据当月的报纸和当月收到的期刊中的重要资料编辑出版的。该索引收录国内公开和内部发行的全国性、专业性的报纸149种，中文期刊达3000余种，成为报导和检索国内中文报纸和期刊论文的重要工具。

《全国报刊索引》（哲社版）采用分类编排，共分11个大类：马列、毛泽东思想；哲学；社会科学总论；政治；法律；军事；经济；科学文化；语言文学；文学；艺术；历史；地理等，大类以下再分为二级类、三级类。在同一类资料中，编排原则采用先动态资料后论文资料；动态资料中又采用先国内后国外；国内动态资料先中央后地方编排。

科技版的《全国报刊索引》报道了国内发行的3193种中央期刊和3种报纸上发表的科技资料，也采用分类编排，它包括10个大类：自然科学总论；数理科学和化学；天文学；地球科学；生物科学；医药卫生；农业科学；工业技术；交通运输；航空航天；环境科学等，大类下再分二级类、三级类。它的著录项目有文献篇名、著者姓名、文献出处、出版年份、期号、页码等。

《全国报刊索引》全面系统地报导了国内报刊文献，为科技人员和图书情报人员掌握国内文献提供了方便的条件。但它所提供的检索途径却只限于分类的一种，尚缺辅检工具，是其不足之处。

该索引每年1月和7月附有“引文报刊一览表”，可以查找文献的出处。

## 二、《中文科技资料目录》

《中文科技资料目录》是在1977年石家庄“全国科技情报检索刊物协作会议”的推动下，通过统一协调和有关部门的努力，由一些专业情报所联合编辑出版的、报道国内科技文献的一套题录性刊物，最多时达到22个分册（详见表2）。

表2 《中文科技资料目录》22个分册

分册名称	编辑单位	创刊年	刊期
1. 综合科技、基础科学	中国科技情报所	1978	6
2. 测绘学	陕西省测绘科学研究所	1978	4
3. 地质	地质部全国地质图书馆	1978	6
4. 医学	中国医学科学院医学情报所	1978	6
5. 中草药	湖南医学工业研究所	1978	4
6. 农业	中国农业科学院情报研究所	1978	4
7. 农业机械	中国农业机械化科学研究所	1978	4
8. 林业	中国林业科学院情报所	1979	4
9. 矿业	辽宁省科技情报所	1978	4
10. 冶金	湖北省科技情报所	1978	6
11. 机械、仪表	四川省科技情报所	1978	6
12. 电力、电工、原子能	陕西省科技情报所	1978	4
13. 电子技术	四机部情报所	1978	12
14. 化学工业	吉林省科技情报所	1978	6
15. 轻工、纺织	天津市科技情报所	1978	4
16. 建筑工程	中国建筑科学研究院情报所	1978	6
17. 建筑材料	国家建筑材料工业局	1978	4
18. 水利水电	水电部科技情报所	1978	4
19. 铁路	铁道部科技情报所	1978	6
20. 公路、水路运输	交通部科技情报所	1978	4
21. 船舶工程	船舶工程索引编辑部	1979	4
22. 环境科学	中国科技情报所	1978	6

《中文科技资料目录》各分册主要报道各编辑单位收藏的文献，报道形式采用题录、简介、文摘相结合，以题录为主。其报道的文献以国内期刊论文为主，兼有少量会议论文、专利、标准。各分册基本按《中国图书资料法》的分类编排，采用国家标准著录格式。各分册在每年第一期刊登“引用期刊一览表”。有的分册出版年度主题索引和其它辅助索引。

该套目录的出版，初步改变了国内科技文献管理混乱、使用不便的落后状况，对充分发挥中文科技文献的作用作出了重要贡献。

### 三、《国外科技资料馆藏目录》

它是由中国科技情报所编辑出版的定期报道该所新入藏的各国政府报告、会议资料、学位论文、研究机构与学术团体著作等国外科技信息资料(特种文献)的检索刊物，共分12个分册。各分册的正文部分按分类编排，著录的内容包括：中文题名、原文题名、文献出处、出版年份、页码、文种等。

它按学科编辑出版，共12个分册，详见表3所示：

表3 《国外科技资料馆藏目录》 12分册

分册名称	刊期
1.数学、力学	6
4.物理学	6
3.地球科学、天文学	6
4.生物学	6
5.医学	12
6.农业、林业、生产	12
7.矿业、冶金	6
8.机制、动力工程、电工	6
9.无线电与自动化	6
10.化学化工、石油	12
11.轻工、纺织	4
14.环境污染与保护	6

#### 四、《国外科技资料目录》

该目录是1977年全国科技情报检索刊物协作会议统一组织协调,由全国各大专业研究所、情报所及专业图书馆分别编辑出版的一套题录性刊物,是查检国外期刊论文与其它文献类型的重要检索工具。目前有36个分册。

《国外科技资料目录》各分册主要报道国外各学科专业的期刊论文、学会出版物、科技报告等文献类型,报道方式一般采用题录,部分辅以文摘形式。各分册的编排方式大都采用《中国图书资料分类法》的分类体系。

下面是《国外科技资料目录》36个分册的详细列表:

表4 《国外科技资料目录》 36个分册

分册名称	编辑单位	创刊年	刊期
1.数学	中国科技情报所重庆分所	1978	12
2.力学	中国科技情报所重庆分所	1978	6
3.激光	上海物理情报研究所	1979	12
4.生物化学、分子生物学	中国科学院上海图书馆	1978	6
5.气象学	中国气象局气象情报研究所	1978	6
6.地理学、地质学	中国科学院地理研究所	1978	6
7.地质学	地质部情报所	1978	6
8.地震学	地震学编辑部	1978	4
9.海洋学	海洋局情报所	1978	12
10.测绘学	国家测绘总局测绘所	1978	4
11.医学	中国医学科学院图书馆	1978	12
12.计量技术	中国计量科学研究院	1974	6
13.矿业	中国科技情报所重庆分所	1978	12
14.机械工程	机械工业部技术情报所	1980	12
15.核科学技术	原子有出版社图书馆	1976	6



(续表)

分册名称	编辑单位	创刊年	刊期
16. 电力	水利电力部科技情报所	1981	6
17. 无线电子学	电子工业部情报所	1974	12
18. 计算机	中国科技情报所重庆分所	1978	6
19. 化学纤维	上海合成纤维研究所	1979	4
20. 日用化学工业	全国日用化学工业科技情报站	1978	4
21. 轻工业	轻工业部科技情报所	1980	4
22. 纺织	上海市纺织科学研究所	1978	6
23. 食品工业	全国食品与发酵工业情报站烟草工业科技	1978	4
24. 制糖工业	全国甘蔗糖业科技情报站甜菜糖业	1978	4
25. 皮革工业	皮革工业分册编辑委员会	1978	4
26. 日用机械工业	全国钟表、缝纫机、自行车工业科技情报站	1978	4
27. 造纸工业	全国造纸工业科技情报站	1978	4
28. 灯泡工业	全国灯泡工业科技情报站	1979	4
29. 建筑、水利工程	中国建筑科学院情报所	1976	12
30. 建筑材料	建筑材料工业技术情报标准研究所	1978	4
31. 水利水电	水利水电部科技情报所	1978	4
32. 铁路	铁道部科技情报所	1978	6
33. 公路运输	交通部科技情报所	1975	12

(续表)

分册名称	编辑单位	创刊年	刊期
34. 水利工程与运输	交通水运规划设计院	1976	6
35. 船舶工程	船舶工程索引编辑部	1976	6
36. 航空与航天	中国科技情报所	1979	12

#### 第四节 国外主要检索工具

##### 一、美国《工程索引》

《工程索引》由美国工程索引公司编辑出版，创刊于1884年，至今已有100多年的历史。目前有年刊、月刊、卡片、缩微胶卷、磁带及光盘等六种版本。它的主要读者对象是工程技术人员、科研人员、工业界及教育界人士等。由于它所报道的文献资料是经过编辑人员精心挑选，并认为有一定参考价值的，所以很受科技界的欢迎。

《工程索引》报道的内容范围较广，但主要是以工程技术各个领域以及科学管理方面的资料为主，纯属基础理论方面的文献一般不予报道。它所报道的全部文献都收藏在美国工程学会图书馆。《工程索引》摘用了世界上48个国家用15种文字报道的科技文献，其中有科技期刊近3500种，以及工程协会、科技学会、大学、试验室和研究所、政府机关和工业企业的出版物、

会议录、特种出版物、研究报告、标准文献、科技图书等，早期还有少量的专利文献。它每年报道的文献侧重面不尽相同，在一定程度上反映出当时的学科的发展趋势及其应用的情况。目前电子、电气、新能源、控制工程、计算机、宇航工程、生物工程、环境工程等学科的文摘占有相当大的比例。

《工程索引》的文摘部分是按照主题词字顺排列，每个主题词后都列出有关的文献，因此，它的文摘部分就相当于主题索引。主题词分为两级，即主标题和子标题，均选自《工程标题表》。1987年以前，在年刊中有“著者索引”、“主题索引”和“著者所在单位索引”。自1987年起，在月刊上增加了“著者索引”和“主题索引”。

## 二、美国《化学文摘》

《化学文摘》由美国化学学会化学文摘服务社编辑出版。1907年创刊，周刊。每年出版两卷，共52期。自1967年起，《化学文摘》报道的内容分为五个方面：生物化学（1~20类）；有机化学（21~34类）；高分子化学（35~46类）；应用化学和化学工程（47~64类）；物理化学和生物化学（65~74类）。

《化学文摘》的特点是：

1. 摘录文献的范围广泛、全面、系统。由于它的历史悠久，实力雄厚，是目前世界各国有关化学化工方面的文献最系统、最全面的检索刊物。它引用了150个国家和地区用56种文字出版的一千多种期刊及其他出版物；26个国家和2个国际组织的专刊。据称其覆盖面已达98%，年报道量已超过60万篇，累计报道的文献已达1000多万篇。

2. 报道速度快。美国书刊一般在发行后12小时内即送到文摘社，报道时差在4个月之内，编辑过程已实现了自动化。

3. 使用方便。各种文字的文摘全都译成了英文，而且各国的著者及所属单位出处，音译成拉丁字母并进行了排序。此外，还提供了多途径的检索入口，编有多种索引，如关键词索引、作者索引、主题索引、分子式索引、专利索引等14种，成为世界上具有比较齐全的检索手段的文摘杂志。另外，它还提供了各种卷索引，5年、10年的累积索引等，因此，查找非常方便。

## 三、美国的《生物学文摘》

《生物学文摘》是由美国生物学文摘社编辑出版的。其前身为《细菌学文摘》（1917~1925年）和《植物学文摘》（1918~1926年），1926年两刊合并后改为现名。

《生物学文摘》是目前世界上收录较丰富、影响较大的有关生物学、医学和农业方面的大型检索刊物之一，其摘录的范围包括世界上110个国家和地区出版的期刊8000多种，包括期刊论文、专题论文、技术报告、学位论文和图书等，年报道量达14万条。自创刊以来，它的总报道量已超过3000万条。它的文摘按主题类目编排，每期有著者、生物系统、属类、概念、主题五种索引。另外还有半年度累积索引。

《生物学文摘》报道的内容共分84个大类，470个小类，包括了生物学中的各个学科，是世界上比较著名的文摘。

## 四、英国《科学文摘》

《科学文摘》由英国电气工程师协会编辑出版，1898年创刊。该文摘共分为三辑：A辑为《物理学文摘》（半月刊）；B辑为《电气与电子学文摘》（月刊）；C辑为《计算机与控制文摘》（月刊）。目前，除有书本、胶卷

与胶片外，还出版了计算机检索磁带和光盘。

《科学文摘》各辑所报道的文献主要是英国电子工程师协会下设的物理、电工和计算机与控制情报服务部（INSPEC）所收集的文献，包括世界上50余个国家以多种文字出版的3000多种期刊、800多个会议的论文集及其他出版物；所收录的文献类型包括各种期刊论文、科技报告、会议文献、科技图书、学位论文以及英美两国的专利说明书（1977年以后不收录专利）；所覆盖的学科主要包括：物理学（A辑）——理论物理、核物理、光、声、电及各种交叉物理学科；电工技术和电子学（B辑）——电磁材料、电讯技术、电力工程、电光学等；计算机和控制（C辑）——计算机理论、计算机技术、计算机应用技术等。

该文摘的各辑都以五种形式出版：书本式、缩微胶卷、胶片和磁带、光盘。由于它每期都附有主题索引、著者索引、参考书目索引、会议索引、图书索引，另有半年和多年累积索引，故查找方便。

#### 五、法国《文摘通报》

《文摘通报》由法国国立科学研究中心的文献中心编辑出版，创刊于1939年，其前身为《分析通报》，是世界上著名的综合性检索刊物之一。

自1947年起它分三个分册出版：第一分册包括数学、物理、化学和工程技术；第二分册包括生物科学、医学、农业及食品工业等学科；第三分册包括哲学、社会科学和人文科学。到1982年，共分为51个分册出版，它全部或部分地收录了世界各国出版的9000多种期刊（占收录文献总量的90%以上），以及会议文献、科技报告、学位论文、专利和图书等，其中英文文献占63%，法文文献占12%，俄文文献占10%，德文文献占8%，其他文种7%。并且年报道量逐年增加，截至1978年，共报道了800余万条。

《文摘通报》收录的文献比较齐全，涉及自然科学和社会科学，各分册的内容均按分类编排。其辅助索引体系比较完善，每期文摘都有法、英两种文字的分类表，且绝大多数分册都附有法、英两种文字的主题索引和著者索引，不少分册还附有专检性索引，每年各分册还出版单独的年度累积索引。其分类目项和各种辅助索引均为检索文献提供了方便的检索途径。

#### 六、美国《科学引文索引》

《科学引文索引》创刊于1961年，是美国费城的科学情报研究所编辑出版的一种多学科文献检索刊物，初始为年刊，自1979年起为双月刊。

由于它为读者提供了一种从参考文献（引文）入手的独特检索途径，加上报道的内容涉及科学技术上许多学科与领域，且选材精练，使用方便，深受世界各国学术界及情报界的重视。

《科学引文索引》所摘引的文献起初包括期刊、专著丛书、图书、会议文献等多种类型。自1981年起，建立了《科技会议录与图书索引》等数据库，为用户提供了联机检索服务，因此1982年后，它只摘引期刊与某些专著丛书。其选材是在对各种期刊进行论文发表数与每篇论文被引用率的统计分析基础之上，共选择了40余个国家出版的质量最优的3000多种科技期刊作为摘引刊物，同时还摘引了200多种重要的科技专著丛书。

该索引报道文献的内容涉及100多个学科或专业，它们分别属于：农业、生物与环境科学、工程、技术与应用科学、医学与生命科学、物理与化学和行为科学等，读者使用方便，每期提供引文索引、来源索引和轮排主题索引。此外还出年度累积版和五年累积版等。

## 七、日本《科学技术文献速报》

日本《科学技术文献速报》由日本科学技术情报中心编辑出版，创刊于1958年。当年出版5个分册，经过若干次调整，目前共出版12个分册，并采用不同的颜色加以区别，各分册日文名称下均有英文名称。

《科学技术文献速报》各分册以期刊形式定期出版，各个分册每期都有卷末索引，包括主题索引、作者索引、报告号索引等。由于它收录的文献较全、报道速度较快、检索比较方便，现已成为与前苏联《文摘杂志》、法国《文摘通报》并列的世界三大综合性检索工具之一。

其12个分册的内容包括：物理与应用物理；化学与化学工业（外国）；化学与化学工业（国内）；金属工程、矿山工程与地球科学；机械工程；电气工程；土木、建筑工程；能源；原子能工程；环境公害；管理与系统技术；生命科学。

《科学技术文献速报》收录世界上54个国家用20多种文字出版的期刊9700种（日本3900种、国外5800种）、16种检索期刊和8种文献数据库（日本3种、国外5种）。它收录期刊论文、各种技术报告、会议资料等连续出版物和部分外国检索期刊中的有关内容。从1976年起，不再报道专利文献。

《科学技术文献速报》选题标准是：凡有科学技术内容的原始文献。

有关科学技术及新产品、新装置等方面的叙述或评论原则上均收录，但对一般入门性文章不予收摘。收录可用于工程设计、产品检查等实用性技术资料，不摘单纯数据、图表和标准方面的文章。其它有价值的一般文章。

它报道范围主要是理工科方面，也报道药学、环境公害、管理与系统技术、生命科学等方面文献，但不报道天文学和基础数学方面的文献。

目前，《科学技术文献速报》载体形式有书本型、标准卡片型、缩微胶片、机读磁带等。

它采用题录、简介、摘要三种方式，以计算机编制，速度快、报道时差约1.5个月。其文献摘要较短，一般不超过300字。

## 八、俄罗斯《文摘杂志》（前苏联）

《文摘杂志》现由全俄科学技术信息研究所编辑出版，创刊于1953年。它引用了世界上130个国家用66种文字出版的2200多种期刊、1900多种图书及26万件专利文献，以及会议录、科技报告、标准文献等。内容包括除医学、农业、建筑以外的所有自然科学、技术科学和工业经济等方面的文献，年报道量达10多万条。1993年出版有27个综合册（含206个分册）和47个单卷本文摘，文摘正文均按分类编排。每期文摘均附有标题分类表，每年第一期的前面有文摘中所使用的缩写字表及引用期刊的刊名缩写和全称的对照表。所有的综合本和单卷本文摘都有年度主题索引和作者索引。另外，根据学科特点，有些文摘还有专用索引，如专利索引、分类主题索引、分子式索引、动植物拉丁文名称索引等，检索方便，所以成为世界上公认的三大综合性检索工具之一。

## 九、其他国外检索工具

除上述八种外，国外的主要检索工具还有：

医学：美国的《医学索引》、荷兰《医学文摘》、前苏联《医学文摘杂志》、日本《医学中央杂志》等。

科技报告：美国《美国政府报告与索引》、《能源研究文摘》、《核子科学文摘》、《原子索引》等。

专利：美国《专利公告》、英国《世界专利索引》、瑞士《专利说明书》、日本《日本专利分类表》、《国际专利分类表》等。

标准文献：《国际标准目录》、《美国国家标准目录》、《英国国家标准目录》、《日本工业标准总目录》、《法国标准目录》等。

会议论文：美国《世界会议》、《会议论文索引》、《科学技术会议录索引》、《会议录出版指南》等。

学位论文：美国《国际学位论文文摘》、《学位论文综合索引：1861~1972》、英国《英国和爱尔兰大学学位论文索引》、《英国学位论文索引》、法国《大学学位论文目录》等。

另外，还有一些比较著名的专业类文摘检索刊物，如《石油文摘》、《农业工程文摘》、《地学文摘》、《地球物理文摘》、《世界铝文摘》、《食品科学文摘》、《海洋学文摘》、《污染文摘》、《分析文摘》、《金属文摘与金属文摘索引》、《化学工业札记》等，这些检索工具都是各个专业领域最常用的高质量文献信息检索刊物，它们中多数已有相应的机读数据库，有的还制成了光盘数据库，以供全球性联机检索服务或在世界各地发行，方便用户检索利用。

## 第六章 参考工具书

文献信息检索工具主要有两种类型，一种类型是用于提供文献信息的线索，从而指引用户根据该线索去查找原始文献，这类检索工具称之为线索型检索工具（可参阅第四章与第五章）；另一种检索工具可以向用户直接提供检索所需的数据、事实或基本概念、基本情况，并被作为获取知识的专用工具使用，这类检索工具习惯上被称为参考工具书，或者称之为事实数据型信息检索工具。

参考工具书属于知识性的检索工具，多以图书的形式正式印刷出版。常用的参考工具书有：字典、词典、百科全书、专科全书、手册、年鉴、名录、图谱、表谱、公式表等。

### 第一节 参考工具书概述

#### 一、参考工具书的特征

参考工具书是广泛汇集某一范围的知识或资料，按特定体例或方式编排，专供查阅的一种特殊类型的图书。

与一般图书相比，参考工具书有其自身的特点。

在内容上，参考工具书既广采博收，又高度浓缩，能提供完整详尽、系统概括的基本知识和高密度的信息资料，具有概括性和知识性。

在编排方法上，参考工具书以科学的编排形式，严谨的结构体系，覆盖有关的知识领域，并提供多种检索途径，具有检索性和查考性。

在外观形式上，参考工具书一般采用精装样式，装帧美观大方，纸质良好，版面清晰，图像逼真，使用寿命较长，便于保存收藏。

在体例结构上，参考工具书有比较稳定的构架。它大体上由序跋、凡例、正文、辅助索引和附录补遗等几部分组成。序和跋亦即前言和后记，一般说明工具书的编纂宗旨、编纂经过、收录范围、内容特点、使用价值等。凡例或说明，主要介绍其编纂体例、编排方法及细则、特定符号等。正文是工具书的核心部分，是查阅的主要内容，能够提供某一领域或综合性的知识信息。辅助索引是供查阅正文部分的各种索引，能提供多种有效的检索途径，辅助索引越多，检索途径就越广，检索效率也就越高。附录补遗是附于工具书正文之后的有关文章、图表、资料、索引等，或补充正文遗漏处及须订正的事项，有助于查考或理解正文，从而达到更好的检索效果。

#### 二、参考工具书的编排方法

参考工具书的编纂目的在于供人查阅使用，属于查考类用书，因此它必须采取一定的方式科学合理地编排所收集的知识单元和信息资料，既能保证使众多的知识单元具有简便的序化，又能不失其间的内在联系，把全部的有关知识充分地表露出来。

参考工具书的编排方式主要有以下几种。

##### （一）字序法

字序法是按照文字所具有的一定的顺序编排工具书词目的方法。它一般用于内容丰富、条目众多的综合性大型工具书，如百科全书、大型字典、词典、名录及指南等。中文的字序有音、形之别。音序法即按字音顺序排列汉字的方法；形序法则是根据汉字形体结构的特点，加以排列的方法。西、俄

文拼音文字多采用字母的顺序，日文多采用五十音顺法。如《新华字典》采用汉语拼音对词条排序，属于字序法中的音序法。为弥补使用上的不便，它在音序排定所收选的 8500 个字之后，又以字形法相辅，在卷首设“部首索引表”和“四角号码表”，作为辅助检索途径，以备查用。

### （二）分类方法

主要是根据参考工具书收录内容的学科性质，兼顾诸多知识单元间的相互关系，逐级细分出不同的类目。它是根据文献资料分类系统来进行编排的。分类方法着眼于知识的全面布局，而各专业分类亦从其整个领域的布局出发，将其专业范围内的知识逐渐细分，显示出上下左右的关系，从而实现族性检索目的。

### （三）主题方法

主题法是将有关资料汇集起来，用能表达和描述文献主题的名词术语来进行编排的方法。它将不同学科领域论述和研究的同一问题或同一事物的文献资料集中于同一主题之下，按字顺加以排列。在参考工具书中一般运用未经规范化的自然语言作为主题词来编排主题资料，不需借助主题词表，以方便使用。用这种方法编排的参考工具书较多，如《毛泽东、周恩来、刘少奇、朱德、邓小平著作主题集成》等。

### （四）主题—分类方法

此法是主题与分类法的一种结合方式，即先按大类集中，再以主题细排，或主题与类目混排，达到从主题和分类两种角度都能进行查检。例如，我国出版的《中国大百科全书》各分册的划分及各分册内的条目安排，就是采用此法的一个实例。主题与类目混排的方法则多用于一些辅助性的工具或专业性强的参考工具书。

### （四）时序法

时序法是按时间顺序排列文献资料的方法。这类文献资料一般包括文献的写作、发表和出版年代，某一具体事件的产生、发展，人生的生卒年及生平活动，以及不同历法的年月日之间的关系。按这种方法编排的参考工具书主要是年表、历表和专门性表谱等。如《中外历史年表》、《中华人民共和国大事记》等。时序法的特点是线索清晰、检索方便。

### （五）地序法

地序法是根据所查资料内容地理位置的顺序或地区次序进行编排的方法。按此法编排的参考工具书主要是地图和地方文献。如《中华人民共和国地图集》等。

按地序法查检，只需根据某一地名或某一资料所在的地区，即能准确查到所需资料。如不知地名和资料所属的确切区域，可利用辅助索引配合查找。

## 第二节 参考工具书的类型

参考工具书的种类繁多，数量庞大，划分的标准也不尽相同。如有综合性参考工具书和专业性参考工具书之分；有社会科学参考工具书和自然科学参考工具书之分；等等。现根据参考工具书的性质和功用标准，可把参考工具书划分为以下几种类型。

### 一、语词性参考工具书

这类工具书主要提供字、词的形、音、义和使用方法，以及学科名词术

语的含义、演变和发展。

语词性参考工具书包括：字典和词（辞典）、百科全书和专科全书。

### （一）字典与词典

字典和词典是字或词的总汇，分别对字或词作出书写方法、读音、释义、用法或概念有关的说明和解释，并按照一定的规则排序，形成各种可供查考的统一体系。字典和词典的主要区别在于收选对象的不同，字典以收字为主，说明字的形体、读音、意义和用法，也兼顾有关的词语；词典则以收词为主，解释词汇的概念、含义和用法，并不都涉及到字。

### （二）百科全书与专科全书

百科全书是以词典的形式编排，系统而又客观地简明地叙述各学科专门术语、重要名词，注重反映学科新研究的大型综合性工具书。

百科全书对各学科及其分支领域所含知识的定义、概念、原理、历史和现状等以条目的方式作出符合客观情况的解释和阐述。因百科全书包括的知识范围广，汇集的知识精，而且主要是作为求知的工具查考使用的，所以，它的编辑体例必须具备概述性、系统性和查检性。

专科全书，也称专科性百科全书，是专科领域里的百科全书性工具，能提供比百科全书更为专深和更为详尽的系统性知识。故而专科全书对其专业的读者的参考作用最大，受到科技界和教育界的普遍关注。许多专科全书不仅提供其专业领域的基本的、系统性知识，而且在专条之后还附有精选的参考资料，以供读者做进一步的研究和查阅。

## 二、资料性参考工具书

资料性参考工具书主要是提供各种基本知识或某一专题的资料。这类工具书包括年鉴、手册、名录等。

### （一）年鉴

年鉴是汇编一年内的重要资料，如学术动态、重要事件、统计资料等，提供事实数据和信息资料，按年度出版的工具书。它多以单卷形式成书，并标以年份出版。其内容丰富，题材新颖，资料性强，篇幅适中，并附有大量的统计数字和图片、表格。

年鉴分综合性年鉴和专科年鉴两种类型。综合性年鉴一般报道前一年间国内外的大事，重要人物的社会活动，一些重要组织机构及其活动的概况，以及重要的统计数字。专业性年鉴多由学术团体、研究机构或政府部门主持出版，报道一年间其专业领域的大事和取得的重要进展。

### （二）手册

手册也称指南、便览、大全等，主要汇集一定范围或某一专业领域内需要经常查询的专业知识，基本情况，常用的公式、数据、规范及规章制度等，以供科研人员放置手边随时随地参考查阅。

手册在内容上有两个特点：一是所收资料侧重基础知识，通常偏重于已成为现实的、成功的具体专业知识，而不是定义、概念、历史的叙述和当前的发展状况；二是实用性强，手册专门汇集人们经常需要查考的文献资料，如各种事实和数据，具有较强的指导意义。

手册的主要作用是能简明扼要地为人们提供各学科专业基础知识和各行业知识。它灵活多样、资料稳定、实践性强，所以是工作学习中不可缺少的方便实用的工具书。

### （三）名录



名录是将机构名、人名、地名以及其他事物名称汇集在一起，按分类或字顺加以排列，并对相关事项予以简要揭示和介绍的工具书。主要包括人名录、地名录和机构名录。

1.人名录。汇集人的本名和别名，并对人物予以简要介绍的工具书，也称名人录。它的收选对象是一定地域、一定时间范围或一定专业领域的知名人物的姓名和简历，以及其在事业方面的成就等。一般人名录比人名字典中的介绍要详细、深入，但不如传记详细、全面。人名录可分为综合性人名录、专门性人名录、今人名录、已故人名录和回溯性人名录等。

2.地名录。著录地名及相关资料的工具书，它可提供地名的标准名称、汉语拼音、类别、经纬度，有的地名录还简要说明地名变迁、人口状况、特殊记事等。

地名录与地名词典虽同是查考研究地名的工具书，但内容和作用不完全相同。一般说来，地名词典对地名的解释较为详细，包括地理特征、历史沿革、风物特产、人口状况等；而地名录则著录内容简明扼要，对地名的主要特征予以揭示，且收录地名详尽，具有规范性和法定性。

3.机构名录。也称机构名录指南，是汇集机构实体的名称并作概要介绍的工具书。其内容一般列出机构的全称、成立时间、组织概况、业务活动、人员配备和负责人姓名等，大都由机构实体、出版单位以及人事部门协作编辑。机构名录又分为国家政府机构名录、经济机构名录、文化机构名录、教育机构名录等。

### 三、表谱性工具书

表谱性工具书是按对象的类别或系统，以表格或其他较为整齐的形式，附以简略的文字，着重反映史实和时间。主要包括年表、历表和专门性历史表谱。

年表按年代顺序编制，专供查考历史年代、历史纪元及历史大事。一般又分纪元年表和纪事年表。纪元年表以时间为主，不述史事，主要用于查考历史年代和帝王庙号、谥号、年号、干支、公元等各种历史纪元。纪事年表以年月为纲，兼载历史大事，供查考历史事件的原委及线索。

历表是将多种计时方法和符号，准确而规律地统一在一定形式的表格上，除可查考历史年代和历史纪元外，主要用于查考换算不同历法的年月日。

专业性表谱包括年谱、人物生卒年表、职官表、地理沿革表等。年谱是逐年逐月记载某人物一生事迹的一种传记著作，也可作为查考历史人物生平的工具书使用。生卒年表将各代名人的字号、别名、籍贯、生年、卒年等项，按人物的生年或卒年先后次序编排，供查考历史人物简历、出生和亡故年代。职官表简要地反映各种职官的名称、建置、职掌的变迁、品级、员额的增减等情况，专供查考我国古代职官制度。地理沿革表主要用于考证我国历代疆域分合、地名演变等情况。

### 四、图录性工具书

这类工具书主要反映各种事物、人物的空间概念和形象概念。包括地图、历史图录。

#### (一) 地图

地图是将地球的部分或全部缩绘在平面上，确切地说，地图是按照一定的数学法则，将地球表面事物和现象缩小、综合，运用符号系统表示在平面上的图像，以及反映各种自然现象和社会景象的地理分布及其在空间与时间

上的相互制约、内在联系和发展动态。

地图按其内容可划分为普通地图和专题地图。普通地图即综合性地图，它较全面地反映地面上事物和现象的一般要素，包括各种自然地理要素和社会经济现象。专题地图突出而深入地表示一种或几种要素和现象，又可进一步分为自然现象地图和社会现象地图，如地质图、水文图、气候图、海洋图等即属自然现象地图，而人口分布图、经济地图、交通运输图、历史地图则属社会现象地图。

## （二）历史图录

它是图绘、摄影等手段来反映历史人物、历史事物的形象的。它汇集了历史著名人物、重大事件、珍贵文献、文化遗址、名胜建筑、书法绘画等图像资料，并配有简略的文字说明，生动地反映政治、经济以及意识形态等方面的变化和发展，是研究历史和文化的形象资料。

历史图录按其内容范围可分为综合性历史图录和专题性历史图录。专题历史图录又可分为文献图录、人物图录、艺术图录、植物图录等。

## 五、公式、数据和规格等实用性工具书

此类工具书多来自有关的专业手册等，是现代科技发展的产物。它除要求尽快修订及时补充新内容外，对技术性的要求更高、更严格，必须合乎技术标准。一些定律、公式、设计要点、符号、代码以及一些特性数据等是其最基本的参考资料，可以说是一种知识信息更为浓缩和密集的参考工具书。这类工具书首先出现于基础学科和技术性的领域，以后发展到其他采用新手段的边缘性领域。如对某学科的数学用表、参数、常数、因数、计算公式等加以集中汇编，作为特殊的新工具出版。

## 第三节 参考工具书使用法

参考工具书种类繁多，质量也参差不齐，每一种工具书都会有其自身的特点和特有的功能，而且具有不同的收选内容、编排结构和查检方法，因此，学会科学使用参考工具书是提高查检效果，提高工具书效用的根本保证。

一般说来，使用参考工具书要注意以下几个方面：

第一必须熟悉各类工具书，包括工具书的类别和用途、共性和个性特点，了解它们的内容特征，然后针对所要解决的问题，选择合适的工具书，找出查检途径，从提供的途径进入正文，对号入座，做到有的放矢。

第二要勤于使用工具书，通过亲身实践，去接触、观察和剖析它们，了解其收选对象、范围、编排体例结构、辅助工具等，不断积累经验，做到熟能生巧。

第三要善于使用工具书。具体到查检一个问题时，一般先从序言入手，再通过“范例”或“说明”及“目次”，查看其收选内容及组织编排结构情况，然后选择最适合的途径查找，才能做到查找迅速，内容准确。

第四要开阔思路，使用多种工具书。在解决重要问题时，不要只查一种或几种工具书就忙于做结论，要多查几种，进行比较、印证。使用时必须互相参照，做到知己知彼。

现从一般需要出发，对工具书的基本使用方法加以简单介绍，并列举各种类型的常用的主要参考工具书，以使读者对工具书使用法和各种工具书有个大致了解。

## 一、字和词的查找

查字和词主要使用字典、词典或辞典。

查字的目的是了解或核实其正确书写方法、读音或确切的释义。此类问题可根据其是常用字还是冷僻字而选择字典。字典的编排多按字音、字形、部首笔画或字顺排，易于掌握。

查词，尤其是查科技词汇，应该使用综合性的科技字典、专业词汇或名称、术语、缩略语词典。词典的排词方法基本同字典的编排相同，外文的编排一般按外文字母的顺序排列。

常见的有代表性的字、词典，简介如下：

### （一）《汉语大字典》

该字典由徐中舒主编，于省吾、王力、吕淑湘等 17 位国内语言文字学专家任学术顾问，四川辞书出版社、湖北辞书出版社 1986~1990 年出版。

全书共分八卷，收单字 56000 多个，是我国到目前为止收汉字最多、单字注音释义最全的一部巨型字典，所有古今文献中出版的汉字，几乎都可以在书中查到。

### （二）《康熙字典》

《康熙字典》是我国第一部以字典命名的工具书，由张玉书、陈廷敬等 30 人奉康熙皇帝之命，在明代梅膺祚《字汇》和张自烈《正字通》两书的基础上加以增补修订，于 1716 年编纂完成，后多次重印。它是古代收字最多的一部字典，共收字 47035 个，按 214 部首编排。全书共分 12 卷，从子集到亥集，按地支顺序排列，每集又分上、中、下三卷。书后有“补遗”一卷，收冷僻字；“备考”一卷，收不通用的字。该字典收字丰富、释义广泛、引证详尽、注音全面。但存在不少错误和缺点，使用时应予以注意。

### （三）《辞海》

该书由《辞海》编辑委员会编，上海辞书出版社 1979 年出版。它的内容涉及 120 余个学科领域，是一部兼有语文和百科性质的综合性词典。全部字头按 250 部首排列，收单字 14872 个、词目 106578 条，附录 13 种很有参考价值的词表。《辞典》有 3 种版本：3 册本、20 分册本和缩印本。其所收辞目由普通词语和专科词语两大部分组成，对辞目的解释简明扼要，主要介绍基本知识，所引资料较为准确可靠，力求反映当代的科学文化水平。

### （四）《辞源》

《辞源》始由陈尔奎等主编，1915 年商务印书馆初版，以语词为主，兼收百科，新旧并容，重在溯源，是查古代汉语词汇的重要辞书。共收单字 13000 多个，词目 10 余万条。后经多次修订改编。1958 年开始重新修订，与《辞海》、《现代汉语词典》明确分工，删去原书中的百科词语，收词下限止于鸦片战争（1840 年），专收古代汉语词汇和成语，以及关于文史学科的条目，成为一部最大的阅读古典文献和研究文史用的古汉语辞书。

《辞源》修订本 1979~1983 年由商务印书馆出版，全书四册，收单字 12890 个，复词共 84134 条，仍按 214 个部首排列，字体也仍用繁体字，字后用汉语拼音注音。1989 年，又出版了《辞源》缩印合订本。

### （五）《现代汉语辞典》

该词典由中国社会科学院语言研究词典编辑室编，商务印书馆 1983 年出版，共 1581 页，是我国知名度最高的语文工具书。它是一部以收录普通话语汇为主的中型词典，所收条目，包括字、词、词组、俗语、成语等 56000 余

条，同时，也收录了一些常用的方言词语、文言词语，以及惯用的专门术语等。释义以现代汉语为准，注意了辞的发展变化，是一部极为重要的现代汉语语文词典。

#### （六）《中国百科大辞典》

由《中国百科大辞典》编委会编，华夏出版社 1990 年出版，1471 页。本书是我国第一部大型、综合性百科词典，涵盖生物学、人类学、医学等 58 个学科，收词条近 5 万，全书约 880 万字。所有条目平实、稳定，力求客观公正，注意反映当代科技的新成果。书后附有表谱达 50 余种。本书按学科编排，文末附词目汉语拼音索引，查检起来比较方便，是一部较好的工具书。

#### （七）《牛津英语词典》

《牛津英语词典》（第 2 版）由英国语文学会组织筹备编纂，1958 年出版，是一部有极高学术价值和实用价值的英语辞书，共 20 卷，收词条 61.6 万多条，引例 241.2 万多条，并有参见条目约 58 万条。第二版与第一版相比，内容增加了 34%，主要增收了近年来的新词语，更新了内容，使其更具有新义，跟上时代的发展。

#### （八）《韦伯斯特第三版新国际英语词典》

该词典初版于 1909 年，第三版于 1976 年出版，共分 2 卷，是语言学界公认的最有权威的英语语文大词典之一。本词典共收词 45 万余条，收录之词均经过严格筛选。为使释义准确、清晰、全面，该词典使用纯解析式的单一短语释义法，释义部分引用了大量例证，其中大多数摘引自 20 世纪中期的著作。词典收录了 1000 万多条引语，引用了 1.4 万多名作者遣词造句的实例。词典也广收现代科技词汇。

#### （九）《朗曼现代英语词典》

《朗曼现代英语词典》初版于 1968 年，是一本具有百科性质的词典。该词典的特点是：释义清楚，注意口语化；例证多，注明惯用法；广收近 20 年来的新词，特别是科技词语。该词典用国际者标注音，对著名人物、地方及重大事件亦作了介绍，附有插图，是一本较为实用的英语工具书。

### 二、人物和事项的查找

查找人物和事项概况，主要使用人名录、百科全书和年鉴等参考工具书。

查人物的简历、传略，可用人名录、人名字典、百科词典、百科全书、专科全书。如问题专深而又具体，在查找过有关工具书后均未查出，或答案过于简略时，则可根据已知线索在专门传记工具书和科学史中寻查，如仍难满意则可借助于文献检索工具中著者索引或机构名录加以补充。

查事项概况是为了解决知识性的问题，可根据具体情况首先使用便捷的的工具书，如年鉴、手册、名录、便览，如仍感不足，还可再查有关词典、百科词典、百科全书、专科全书或大全类的专业工具。如属于比较专深的问题，可直接查专业手册、专科全书等。

查人物和事项较常用的参考工具书有：

#### （一）《中国大百科全书》

它是我国第一部大型综合性百科全书。由中国大百科全书总编辑委员会编，中国大百科全书出版社 1980 年起陆续出版。按预定计划出版 80 卷，1987 年 9 月总编辑委员会为保证其质量，经批准同意将卷数减为 74 卷（内含索引卷）。除索引卷外，已于 1993 年出齐。该书按学科分类分卷出版，包括 66 个学科知识领域，社会科学与科学技术的比例各占 50%。全书共收条目 77859

个，约 12568 万字。其内容上最大的特点是新、精和实用。它既关注基础，又偏重前沿；既兼顾过去，又重视现代；既侧重中国，又涵盖世界。它阐述的基本知识和提供的学术资料，其广度、深度和质量使之成为一个比较完整的知识大系。

#### （二）《简明不列颠百科全书》

它是中国大百科全书出版社和英国不列颠百科全书公司合作编译，中国大百科全书出版社 1985～1986 年出版。《简明不列颠百科全书》中译本是根据英文版《不列颠百科全书》第 15 版的《百科简编》部分编译而成，基本上保持了原书的全貌，其中有关中国的条目由我国专家学者重新撰写。全书共 10 卷，1～9 卷为全书的正文及附录，第 10 卷为索引，共收选条目 71000 余条，总计约 2400 万字，附有图片 5000 幅左右。其内容包括社会科学、自然科学、工程技术、文学艺术等各个学科的概述和专有名词、术语及世界各国人物、历史、地理、团体、机构等介绍，侧重介绍西方的文化、科技成就和当代知识。该百科全书在原著的基础上，根据我国的实际，经编译和局部修改后，更适合我国读者的使用。其内容简明，查阅简便，编排合理，是一套具有很高的应用价值的工具书。

#### （三）《科学技术百科全书》

它是由中国科学出版社于 1980 年出版，是美国麦格劳—希尔图书公司出版的《科学技术百科全书》（1977 年，第四版）的中译本。它汇集了近代世界基础科学和技术科学的主要成就，是一套多学科的科技工具书。原书由美国、英国、日本等国的科技界名人和专家参与组织编纂，收词条约 7800 篇，内容包括科学技术领域的 100 多个学科。该书在美国出版后，受到国际科学界的重视。中译本的《科学技术百科全书》按学科分卷出版，共 30 卷。前 29 卷为条目正文部分，最后一卷为索引卷。

#### （四）《英国百科全书》

这卷百科全书被认为是现代最有权威的大型综合性百科全书，也是著名的三大百科全书之一。全书 1768～1771 年创编于苏格兰的爱丁堡，至 1974 年刊行了第 15 版，书名为《新不列颠百科全书》。

该版共分 30 卷，收录词目 10.6 万条，全书分成三个部分，各有名称和副题。第一部分称《简编百科》，副题为《便览和索引》，共 10 卷；第二部分称《详编百科》，副题为《知识深义》，共 19 卷；第三部分称《类目百科》，副题为《知识纲要和百科指南》，为一卷。

《简编百科》既是整套百科全书的便览，又是它的索引；《详编百科》是这套百科全书的主体，它包含了按字顺排列的 4207 个大条目，这些条目都是邀请世界各国著名学者撰写的专门论著，对主要学科和重要的人物、事件都有详尽的介绍和叙述；《类目百科》既是一个独立、完整的知识体系概要，又是阅读使用整套全书的指南。

#### （五）《中国百科年鉴》

该书由中国大百科年鉴编辑部编，中国大百科全书出版社出版，1980 年创刊。它是为《中国大百科全书》补充和提供资料而编的，是我国第一部综合性大型百科年鉴。

该年鉴正文前是署名文章，分两种：一是“专文”，着重论述一个领域的形势；一是“特稿”，主要反映党政领导人的重要讲话和文献。其后是“大事记”、“新闻人物”、“诺贝尔奖金获得者”、“逝世人物”几个小栏目。

正文分概况、百科、附录三部分： 国内外基本情况综合； 分类介绍我国政治、经济、科学、文化等诸方面的进展； 附录各种统计资料，书末附有索引。

#### （六）《中国年鉴》

该年鉴由北京新华通讯社《中国年鉴》编辑部编，北京中国年鉴社、香港新中国新闻有限公司出版，1981年创刊。这是一部图文并茂的综合性年鉴，它分中国概况、特载、彩图专辑、大事记要，分类条目和附录等部分，其中分类条目是全书的主要部分，它以部类为单元，共分政治、法制、军事、外交、财政金融、经济、科学技术、文教卫生、体育、社会生活、新闻人物等 11 个部类。检索各学科动态应以这部分入手，使用时可利用书前分类目录。

#### （七）《中国人口年鉴》

该年鉴由《中国人口年鉴》编写组于 1985 年编辑出版。它编入 1954 年以来我国的人口政策；三次人口普查的主要数据；第三次人口普查的详尽分类统计；我国历年人口资料；我国千分之一人口生育率抽样调查材料；中国人口与世界人口对比；各省、自治区、直辖市人口发展综述；有关的调查报告；我国的人口问题专论等。

#### （八）《国际名人录》

该书由英国《国际名人录》编辑部编辑出版，创刊于 1935 年。它是一部查找世界当代知名人士的资料型传记工具书，收录人物全，亦较准确。本书每年出版一次，每版均有增补。此书收录范围广，包括世界各国政治、经济、法律、外交、军事、宗教、文艺、科技界等著名人士。著录项有姓名、生年、国籍、专业、学历、家庭、职务、著作、爱好、通讯处等，按姓氏字顺排列。我们查找当代著名人物应首先查检这本书。

#### （九）《中国当代名人录》

《中国当代名人录》由中外名人研究中心编，上海人民出版社 1991 年出版。该名录收当代中国在政界、军界、实业界、教育界、科技界、文化界等领域有突出贡献者、地位重要者以及知名度较高者 7564 人，所录人物释文内容包括生年、籍贯或出生地、主要经历、主要贡献。所依据的资料绝大部分由本人提供，个别部分则参照相关资料或研究成果而编写。资料征集截止日期为 1990 年 12 月。该书是一部查找我国当代知名人士的综合性工具书。

#### （十）《海外华人名录》

该名录由诸有钧编，中国友谊出版社 1994 年出版。它收录的海外华人包括居住世界各地、在各个领域有优异成就和卓越贡献者，各国各地华侨、华人社团组织之领袖人物，对当地社会有突出贡献的知名华人，在各国（地区）各级政府部门任高职务者，关心中国建设及统一事业的知名人士等。所收人物，由中国大陆和海外知名人士组成的推荐委员会提供入选。

### 三、地名及机构名的查找

查找地名可首先使用地名录、地图集、地理图册、地名词典等参考工具书，有时还可使用百科词典、专业手册来查检，或者直接查大百科全书，如利用其书后的地名索引当更准确，书中对异地同名的地名及历史地名都有特别注释，以示区别。

查找机构名可利用机构名录和各种词典。这种工具覆盖面广，除有科学、行业之分外，还有国际、地区、国家、地方等之分。这类名类自成系列，一

般对本国情况的报道更全更准更新一些，因此，查此类问题应首先考虑该机构所在国出版的工具书。

查找地名及机构名的参考工具书举例如下：

(一) 《世界地图集》

《世界地图集》由地图出版社 1972 年出版，这是我国出版的一部比较详密的世界地图集，共 85 幅。主要分成世界和欧洲；分区分国图；地区补充图三级。每幅图均附有文字说明，简要介绍各国和各地区的概况、自然资源、经济等内容。

(二) 《世界地名录》

《世界地名录》由萧德荣主编，中国大百科全书出版社 1984 年出版。该名录收中外地名近 30 万个，正文分外国地名和中国地名两部分。外国地名部分收录了中国地名委员会编辑的《外国地名译名手册》和《泰晤士世界地图集（1981 年版）》中全部外国地名词条；地图出版社出版的非洲分国地图，美国国家地名局出版的亚洲、美洲分国地名图录和《不列颠百科全书》（简编）中重要地名，也酌量收入。所收外国地名条目一律按罗马字母顺序混合编排，中国地名条目按汉语拼音顺序排列。对条目的解释，外国地名一般包括罗马字母拼写、中文译名、所在地域和地理坐标四项内容；中国地名一般包括汉语拼音、中文地名、地理坐标三项。卷首卷末有附录 10 种。

(三) 《地名翻译手册》

《地名翻译手册》由萧德荣主编，知识出版社 1988 年出版。本书共收录中外地名 30 万条，正文分外国地名、中国地名两部分，并有若干附录以供参阅。该书收录的地名均以西文名称在前，中文译名列后的形式编排，并按西文字母顺序排列。本书除供编纂百科全书审核地名外，对查找地名也有较大的参考价值。

(四) 《中华人民共和国地图集》

该地图集由地图出版社 1979 年出版。本书由专题图、省（区）图、城市图三部分组成，并附有文字说明，同时附有地名索引，是一本介绍我国地理面貌、行政区划和经济建设成就的综合性地图集。

(五) 《中国政府机构名录（1992 年版）》

该名录由新华出版社 1993 年出版，经国务院办公厅秘书局和中央保密委员会同意，由新华社和国家编制委员会办公室共同征集和编辑。全书详录了国务院、各部委、国务院直属机构及办事机构，部委所属司局及归口管理的国家局，以及司局所属的处室；各省、自治区、直辖市、计划单列市及其所属厅局和处室；省辖市、地，直到县属科局。包括这些机构的名称、地址、邮政编码、电话、电报、传真和负责人姓名，以及司（厅）局以上单位的主要职能等，采用材料截止于 1992 年 6 月。全书分为中央卷（一册）和地方卷（八册）。

(六) 《各国国家机构手册》

该手册由新华社国际部国际资料编辑室编，中国对外翻译出版社 1993 年出版。它介绍了 194 个国家和地区的国家机构，其中分国家元首、议会、政府、司法机构、武装部队及政党等项情况。各国的重要机构均加以简要说明，介绍其职能和组成情况。先分洲编排，再按国名的英文字母顺序排列。书后附有各国面积、首都、国庆日、独立日以及同中国建交日期等一览表。

(七) 《中国工商企业名录大全》

该大全由中国国际广播出版社 1992 年出版，共分 8 册，是目前收录商务业务信息最新最完整的大型名录。主要编入截止 1991 年 7 月底之前成立的具有法人资格的我国各行业工商企事业单位的名称、邮政编码、通信地址、电话、电报挂号及部分经营范围、产品简介。该书勾画了现阶段科研、工业、商业、农贸等诸多行业的基本框架，共分 312 大类，1658 小类。

#### （八）《世界工商企业大全》

该大全由中国现代国际关系研究所编，中国友谊出版社 1993 年版。主要介绍美国、加拿大、西欧、日本、韩国、东南亚、独联体、港澳台及第三世界国家著名工商企业的情况，包括经营范围、资产状况、企业主营简历、联系通讯方式和驻华机构等最新信息。

#### 四、年代日期及统计数据的查找

查找年代日期，如属历法换算可查对照性历表，如查时代发生的时间，则可查年度大事记、回溯性大事记、年鉴、综合性手册、百科词典和百科全书等。

查统计资料和相关数据，可利用各种统计资料汇编，也可使用各种统计年鉴以及各专业年鉴中的统计栏目，还可利用一些专门性手册、专业手册、教科书、专业书及专科全书等。

查找这两类问题，可利用以下工具书。

#### （一）《中国历史纪年表》

该年表由万国鼎编，中华书局 1978 年出版，全书分上、下两编。上编包括历史年代总表和公元甲子纪年表；下编包括 12 个表：夏商周年代简表、殷年代简表、西周周王简表、东周周王简表、东周诸侯国年表、西周诸侯存亡表、秦以后主要朝代存亡表、十六国年表、十国年表、中日对照年表、公元甲子检查表、太岁纪年表。附中国历史纪年表索引。公元甲子纪年表是该书的重要部分，包括公元前 841 年至 1949 年间公历与中历的年代对照。

#### （二）《中国史历日和中西历日对照表》

该表由方诗铭等编著，上海辞书出版社 1987 年出版，是一本供换算中西历年月日的参考工具书。该表分上编、下编、附篇三大部分。上编为公元前历日表，起于西周共和元年，迄于西汉哀帝元寿二年；下编为公元后历日表，起于西汉平帝元始元年，迄于中华民国三十八年；附篇列有殷历日表、共和元年前西周历日表、1949~2000 年历日表、韵目代日表、六十干支表等。该表按年编排，书后附年号索引。

#### （三）《新编中国三千年历日检索表（前 1500~2050）》

该表由徐锡祺编，人民教育出版社 1992 年出版。在目前已出版的供查中西回年月日的数种历表中，该表检索时限最长，检索范围最广，推算也较简便，既可解决中、西、回、日、朝、越年代的互换，又可推算中、西、回、和历年月日，并配有各种用途的附表。该表分历日检索表、列国纪年表和参考资料三部分。

#### （四）《两千年中西历对照表》

该表可供换算中历与西历年代，由薛仲三等编，三联书店 1956 年增订重印。该表反映了公元元年至公元 2000 年间的中历、西历的年月日对照。可由已知的公元日期直接查出与其相当的中历日期，并可推知每日是星期几和纪日干支。

#### （五）《中国统计摘要》



该书由国家统计局编，中国统计出版社于 1983 年起逐年出版，以反映上年度国民经济和社会发展为主。1984 ~ 1987 年诸本简要列入了建国以来主要年份和 1978 年以来逐年国民经济各部分的主要统计指标数字。从 1988 年本开始舍去建国以来主要年份的统计数字，各年度本分类情况不一，1994 年本分为综合、人口和劳动力、固定资产投资、财政、物价、外贸和旅游、人民生活、农业、工业、建筑业、运输邮电业、商业、教育、科学文化、卫生体育等部分。

#### （六）《世界经济信息统计汇编》

此汇编由陈秀英编，中国物价出版社 1994 年出版。它汇总了国内外世界经济信息统计的精华，突出反映了 90 年代世界经济发展的现状、趋势及热点问题，分为综合与专题、国际组织与国别两大部分，还收入了台湾、香港、澳门近期的统计资料和国际常用度量衡换算表两个附录。全书共设 120 个专题，包括四个国际组织及世界 40 个国别地区的统计资料。附有主要工业国 1993 ~ 1994 年经济形势短期预测资料。

#### （七）《中国社会统计资料》

该书由国家统计局社会统计司编，中国统计出版社 1985 年起陆续出版。1985 年出版第一本，时间范围大体是从 1949 年至 1983 年。它收录了全国和各省、自治区、直辖市社会发展状况的统计数据，包括自然环境、人口与家庭、劳动、居民收入与消费、住房与生活服务、劳保与福利、教育、科研、卫生、环保、文化、体育、社会秩序和安全、社会参与共涉及 14 个部分的内容。

#### （八）《中国统计年鉴》

《中国统计年鉴》是查检我国各行业统计数据的一部具有权威性的综合性年鉴，由国家统计局编，中国统计出版社出版，1982 年创刊。该年鉴全面反映我国经济和社会发展情况，收录全国和各省、自治区、直辖市经济和社会各方面大量的统计数据。其内容除包括我国各行各业当年各项统计指标和资料外，还另附台湾省主要经济指标，以及我国经济、社会统计指标同世界主要国家和地区的比较。

## 第七章 信息存贮

### 第一节 信息存贮的意义

信息存贮是信息工作的基础，也是科学技术乃至整个社会得以存在和发展的基础。

#### 一、信息存贮是科学知识得以积累与流传的基本条件

人类存贮信息具有三种方式：大脑存贮、实物存贮和文献存贮。长久以来，人类一直在不断地探索和寻求保存信息的方法和载体。结绳刻痕是人类最早的、低级的信息存贮的方法；泥土、石块、甲骨、竹简、丝帛都曾作为信息的主要载体；直到原始社会末期文字的发明，使得人类信息的传递突破了口语的直接传递方式，使信息可以存贮在文字里，从而超越了直接的时空界限，是人类解决信息存贮、信息表达、信息交流、信息载体的一次本质飞跃。仅文字演变就经历了甲骨文、钟鼎文、小篆、隶书、楷书、草书和行书等阶段。造纸和印刷术的发展，再一次把信息的存贮和传递带入了一个新的境界，扩大了信息存贮和交流的容量和范围，使人类文明得以迅速传播。

文献存贮已成为知识存贮的主要手段。大脑存贮和原始的实物存贮的容量与时间都是有限的，不能长期保存、广泛交流和充分利用。而存贮于文献的信息知识既是大量的，又不受时间、地域限制，能够长久地保存传播。因此，只有通过各种载体形式的文献，才能把人类获得的科学知识加以存贮、积累，世代相传，才能促进人类社会的发展。

#### 二、信息存贮是信息工作的重要内容

从早期的文献搜集、整理到今天电子计算机信息数据库的建设，从古代私人藏书、藏书楼到近代图书馆，直到现代化图书信息服务机构的建立，信息存贮一直都是文献信息工作的主要内容。没有文献信息的存贮，就没有人类文明和知识的流传和继承，更谈不上知识体系的完整构建和迅速发展。

信息存贮手段的不断发展和进步，增强了人类加工利用信息的能力。随着现代信息技术，尤其是电子计算机的发明和应用，从本质上改变了信息存贮依赖于文献载体的基本方式，也突破了人类大脑及感觉器官加工处理信息的局限性，使信息存贮进入了一个新的阶段。当今，磁存贮、缩微存贮、半导体存贮、光存贮的先后出现和并存，使信息存贮向着大容量、高密度方向发展，信息存贮工作也逐步迈向机械化、电子化、自动化，极大地提高了信息存贮的效率。

#### 三、信息存贮是整个信息检索工作的基础

信息检索的全过程包括信息的加工存贮和检索利用两个方面的内容。信息的检索与利用依赖于信息的积累和加工存贮，可以说，没有信息的存贮工作，就谈不上信息的检索工作。

信息的存贮过程，主要是实现整个信息检索系统的存贮职能。从文献信息源接收信息，经鉴别、筛选和分析加工，把信息的主题内容转换成主题概念或外形标识（如分类属性、主题、著者、篇名等等），然后进行有序化整理，形成特定的信息集合，以待利用。这种信息的存贮过程是检索系统的基础，同时也是检索系统得以存在的前提条件。信息的检索过程也正是在此基础上加以进行和得以实现。

目前，信息的存贮已由社会上的专门信息机构大规模地经营，其存贮结

果则采用不同方式以各种媒质广泛地提供社会使用或出售，同时，又在内部日积月累，形成各种数据库，并进而同其它数据库联网，广泛提供联机服务。这种规模不等、专业各异的信息加工存贮劳动，一次取得的成果可供更多的人广泛使用，从而使信息的存贮结果实现更大范围的传递和共享。

新型存贮检索技术的发展，为信息的存贮和检索展现了广阔的前景。电子计算机中存贮的信息，也正由早期的数据、文献的著录项、文摘等向全文信息、图像信息、视频数据信息、多媒体信息发展；信息检索的范围也将日益扩大，由文本检索向多媒体检索以及超媒体检索发展。

## 第二节 信息存贮技术

从早期以印刷品为载体存贮信息，到现在以软盘、硬盘、缩微平片和光盘等新型载体存贮信息，可见信息存储技术有了迅速的发展，为信息的存贮提供了更优良、更便利的条件，不仅使信息存贮高密度化，而且使信息存贮与快速检索结合了起来，极大地提高了信息存贮与检索的效果。

下面对信息存贮的主要技术加以介绍。

### 一、信息的印刷存贮

造纸和印刷术的发明，为信息的存贮与交流带来了深刻的影响。

印刷是指将文字、图形等信息经过一定的工艺操作，成批量地复制出来。随着印刷术的日益精湛，在各种类型的印刷载体，如各种纸张、纺织品、皮革、塑料、玻璃、陶瓷上印刷的效果，已经达到了相当精美的程度。

纸张以外的印刷载体，如纺织品等，尽管也起到了存贮、传递、交流信息的作用，但主要还是为生活用品及装饰用品增添艺术色彩，它们并不适合作为积累和保存大量信息的载体。长期以来，世界各国的图书馆、档案馆、文献信息中心、资料室等公益性的文献存贮机构，也正是一直以纸张印刷文献为保存对象，以达到信息存贮、交流、利用和实现文献信息资源共享的目的。

信息的印刷存贮，对人类文明进程起到了巨大的促进作用，为人类知识的积累和传播做出了重大的贡献。但是，信息的印刷存贮技术发展到今天也显露出许多弊端。首先，信息的印刷存贮体积大、容量小、占用空间大，在文献信息日益增多、极度膨胀的今天，各文献收藏单位，如图书馆的建筑面积也远远跟不上文献增长的速度，从而造成文献收藏机构的空间拥挤，难以进行更大范围、更多的文献收藏；第二，信息的印刷存贮速度慢，印刷过程复杂，印刷周期长，造成文献信息出版的“滞后”现象；第三，信息的印刷存贮使文献信息的传递速度和传递范围受到影响，不能快速地实现信息内容的有效传送，更难以达到全球范围的普遍共享；第四，信息的印刷存贮也造成文献管理的困难。由于纸张印刷文献的寿命有限，又容易受到一些外在因素的损害，如温度、湿度、光线、灰尘、蛀虫等都会对纸张的保存产生影响，所以纸张印刷文献的管理要考虑多方面的因素，才能进行妥善管理。

虽然，信息的存贮至今仍离不开印刷存贮，并且印刷存贮也仍然是信息存贮的主要方式，但是，人们正越来越多地采用其它更为先进的信息存贮技术，而且新的信息存贮技术也将逐渐替代传统的印刷存贮。

### 二、信息的磁存贮

在现代信息存储技术中，磁存贮是信息存贮的主要手段，磁存贮信息系

统，尤其是硬磁盘存贮系统，是当今各类计算机系统的最主要的存贮设备。

### （一）磁存贮的特点

磁能存贮一切可以转换成电信号的信息，如声音、图像等。它具有如下特点：

1. 信息能长久保存在磁带中，并可重复使用，而一旦所录信息无用时，又可随时抹去，再重新记录新信息。
2. 能同时进行多路信息的存贮，而且当采用多路频率调制方式进行存贮时，能保证这些信息之间的时间和相位关系。
3. 存贮频带宽，可存贮从直流 2 兆赫以上的信号。
4. 存贮时可根据对象过程和研究的需要，对高频信息采用高速存入、慢速复放的方式；或者对低频信息采用慢速存入、快速复放的方式。

### （二）几种主要磁存贮介质

1. 计算机磁带。磁带是最早出现的一种磁表面存贮载体，它始于录音介质，主要用来记录模拟信号。计算机问世后，美国 IBM 公司完成了将磁带作为计算机存贮信息载体的研究。磁带是磁介质存贮中成本最低但信息存贮速度最慢的一种。

磁带至今仍广为采用。它是由载体层和磁表面记录层组成的。供计算机用的磁带宽度大都为 0.5 英寸，长度有 730 米、385 米、90 米数种，磁带的存贮密度有 800bPi、1600bPi、6250bPi（bPi 表示位/英寸）数种。近年来越来越趋于高密度。

磁带放在磁带机上工作，而磁带机又是在中央处理器（CPU）与磁带控制器的操纵和控制下运转的。使用时通常把若干台磁带机与一台磁带控制器相连，再将磁带控制器与计算机相连。现在，大型计算机系统一般采用真空积带箱式磁带机，它是一种高速、高密度的带盘式磁带机。最近出现的数据流磁带机取消了主动轮与真空积带箱，结构小型简单，可在磁带上高速记录数据，数据存取速度与传输率都较高，成为最理想的温盘后备存贮器。

磁带存贮器的主要优点是价格便宜，存贮量大，占用空间小，性能价格比高。而其缺点是只适于顺序存取，而且存取速度有一定限制。工作时由于磁头要与磁带表面相接触，故容易损坏磁表面层。磁带存贮器的顺序存贮方式决定了它特别适用于大批量的回溯检索。

2. 硬盘。硬盘又称硬磁盘，是在铝合金圆盘上涂有磁表面记录层的磁记录载体，硬盘的直径有 14 英寸、8 英寸、5.25 英寸和 3.5 英寸等多种，其中以 14 英寸的硬盘用得最多。

硬盘通常由多个盘片组成，称为盘组；每张盘片由若干闭合同心圆组成磁道；盘组中同一半径的磁道构成一个立体的筒壁，称为柱面；每面上又被均匀地以扇子形状划为若干段，每段称为一个扇区。不同磁道上的扇区弧长虽不相同，但存贮量相同。一般说来，盘的大小不同、密度不同，盘面上的磁道数也不同，扇区的分法也不尽相同。

磁盘组是由多个固定在同一根轴上的单片硬盘构成的，磁盘组的单片盘数可由几片到几十片，存贮容量从几百兆字节到几千兆字节。磁盘组可在磁盘驱动器的带动下旋转，以便读写数据。磁盘上信息的读写是由计算机的中央处理器通过磁盘控制器来控制的。一台控制器也可与多台磁盘驱动器相连，其作用与磁带控制器相同。

磁盘存贮器的最大优点是能够随机存取所需要的数据，数据传输速度

快，适合作为大容量的检索设备。这些优点使其成为最重要的计算机外在设备。

3. 软盘。软盘技术诞生于本世纪 70 年代。目前，软盘已成为微机系统中不可缺少的一种外存贮介质，而大、中、小型机也将其作为一种主要的信息输入载体。在软盘机中一次可放入 20 张软盘，能连续将数据输入到主机中，大大提高了输入效率。

软盘又称为软磁盘，是在柔性塑料圆盘上涂有磁记录层的载体。软盘的直径有 8 英寸、5.25 英寸、3.5 英寸等几种，其存贮容量大都在 180kB ~ 3MB 之间（kB 表示千字节），标准 5.25 英寸高密度软盘的容量为 1.6MB。新的直径为 3.5 英寸的硬质软盘的存贮容量在 10MB 以上，能很方便地在磁盘机上装卸，存取速度比普通软盘快 20 倍。

软盘的优点是它的驱动器体积小，重量轻，结构简单，价格低；软盘的缺点是存贮容量小，存取速度与数据传输率都较低。目前，在计算机病毒流行的情况下，软盘是病毒的危险载体，使用软盘时应注意防治病毒。

4. 温盘。温盘是温彻斯特磁盘的简称。温盘是本世纪 70 年代中期发展起来的新磁盘技术，目前大多数硬盘装置都采用这种技术。

温盘尺寸有 14 英寸、8 英寸与 5.25 英寸等数种。大容量的 5.25 英寸温盘机的存贮容量已达 770MB。

温盘具有以下基本特征：

- （1）采用组合件方法消除影响磁头定位精度的机械变动因素；
- （2）采用密封的防尘结构减少浮动高度和有效记录磁道宽度；
- （3）采用体积小、重量轻、负荷小的磁头和表面润滑磁盘，使磁头可靠地按接触起停方式在小浮动高度下工作，从而消除磁头集中加载对盘面的冲击可能造成的头盘损伤；
- （4）采用薄的高性能磁盘媒体提高读写性能；
- （5）采用读用集成电路，并尽可能把它安装在靠近磁头处，以改善高频信号的传输质量。

温盘的上述特征增加了系统的稳定性与可靠性，对于提高磁盘机的性能价格比也起到了重要作用。

### 三、信息的缩微存贮

缩微技术是缩微摄影技术的简称。缩微技术的胶片作为介质，采用感光摄影原理，并逐步与计算机、微电子、静电复印与传真等现代先进技术相结合，成为迄今最成熟的文献资料全文本真迹存贮和检索技术。

信息的缩微存贮是用缩微摄影机将文件资料缩小拍摄在感光胶片上，经加工处理后作为信息载体保存起来，供以后拷贝、发行、检索与阅读之用。

缩微技术创造于 19 世纪 30 年代，英国人约翰·丹塞于 1939 年成功地将 20 英寸的文件缩微成 0.13 英寸。但缩微技术真正广泛用来存贮传递信息还是从本世纪 30 年代开始，它以独特的形象逐步在信息存贮与交流中发挥出愈加突出的作用，尤其是 70 年代以后，缩微技术先后与计算机技术、光盘技术相结合，组成了更加完美的信息存贮与检索系统，使这一技术更受重视。

#### （一）缩微存贮技术的特点

1. 缩微品的信息存贮容量大，密度高。
2. 存贮介质占用空间小，可节省大量空间。由于缩微品体积小、重量轻，在存贮相同数量资料的情况下，缩微平片比普通纸节省存贮空间 98%，其体

积与重量仅为印刷品的 1%。

3. 缩微品忠于原件，不易出差错。

4. 成本低、价格便宜。

5. 保存时间长，在通常环境下缩微品可以保护 50 年，如果在适当环境温度和湿度下可以保存 100 年以上。

6. 便于计算机检索。采用缩微技术，可以将非统一规格的原始文件规范化和标准化，从而便于管理。

但是，缩微存贮技术也存在一些缺点，它在检索与阅读时需要专门的缩微阅读器，长时间阅读易于疲劳；不能像在纸印刷品上那样在缩微品上进行批注，造成修改困难。

## （二）缩微制品的分类

缩微品按其外形可分为卷式胶片与片式胶片两大类。卷式胶片采用 16 毫米和 35 毫米的卤化银负片缩微胶卷作为记录介质，胶卷一般长 30.48 ~ 60.96 米。卷式胶片成本低，所需存贮空间小，存贮容量大，安全可靠，但检索使用不便，适于存贮使用率较低的大批量资料。片式胶片又可分为缩微平片、条片、封套片、开窗卡片等。

按材料来分，缩微胶片又可分为卤化银胶片、重氮胶片与微泡胶片三种。卤化银胶片是将含有感光溴化银或氯化银晶粒的乳胶涂在塑料片基上制成的，其最大的优点是保存时间长，但容易受温度和湿度的影响。卤化银胶片是目前使用最广泛的胶片。供用户使用的拷贝片一般采用价格较低的重氮胶片或微泡胶片，它们的缺点是保存时间短。

## （三）缩微技术的现代应用

缩微技术最令人瞩目的进展就是它与电子计算机、通信技术及其他存贮介质的结合，拓宽了它的应用领域，发挥了它的潜能。其主要成就有：

1. 计算机输出缩微胶片 (COM) 技术。COM 技术能将计算机输出的二进制信息转换成人可阅读的缩微影像，并直接将它们记录在缩微片上。COM 的输出速度很快，每秒可摄录数千乃至上万个字符。COM 设备可起到一个档案库存贮器的作用，将机内占较大容量而又需长期保存的信息，利用 COM 输出到缩微胶片不失为一种简便有效的方法。

2. 计算机输入缩微片 (CIM) 技术。CIM 与 COM 的作用正好相反，它是把缩微胶片上的字符、文字、图像等信息转变为电脉冲，以机器可读形式记录在磁介质和光介质载体上。同样，再通过 COM 又可获得原始缩微胶片的复制品。因此，CIM 具有进一步推广 COM，扩大缩微品应用的重要作用。

3. 计算机辅助缩微品检索系统 (CAR)。CAR 是一种将计算机信息检索技术、缩微品及纸质资料各自特点融为一体的自动化检索系统。它的主要设备有：计算机系统、缩微品存取设备、缩微阅读机和缩微复印机。其实现方法为：将缩微胶片信息中的可检索项（主题词、关键词、分类号、代码、日期等）以及信息所在胶片的地址（卷号、片号、顺序号等具有指导意义的信息）输入计算机，并组成索引；通过计算机外设输入检索要求，检索软件逻辑判断，获得命中结果（信息所在缩微品中地址），由此而驱动并控制缩微存取设备取出原文，并显示在缩微阅读器上或由缩微复印机输出纸质原始资料复制品。目前，CAR 已具有在 1 分钟内，从一万页缩微资料中检索出任意一页的能力。

4. 视频缩微系统。它是由缩微、视频和计算机三种技术结合在一起构成

的影像资料全文存贮检索系统。由于采用视频技术，可以做到输入的影像资料先暂时存入磁盘而实现随存随用，克服传统缩微系统因制备缩微品需要时间而不能立即将新资料供用户使用的固有缺点。而且输出的缩微影像经扫描数字化，可以通过线路向远方传送，使网络中众多的用户不但可以同时共享机读二次文献而且还能共享全文影像资料，从中找出任意一页原文献只需 14 秒钟。

总之，缩微技术已经发展成为一种相当成熟的技术，并且已广泛应用于保存珍贵的文献和典籍上，世界各地的大型图书馆常采用该技术对珍本、善本和孤本进行缩微处理，而且还利用计算机辅助缩微品检索系统和视频缩微系统，实现全文检索，并逐步通过通信网络，实现缩微品自动存贮检索和缩微全文资料共享。

### 第三节 新型存贮检索技术

#### 一、信息的光盘存贮

光盘存贮技术是利用激光和计算机存贮信息的最新技术。

信息的光存贮技术起始于 20 世纪 60 年代，70 年代得到迅速发展。早期主要是研制激光式电视唱片。光盘存贮器的出现，是信息存贮技术的重大突破，它为海量信息存贮和信息检索等应用领域提供了广阔前景。

光盘是用聚焦成直径小于 1 微米的激光束在光记录介质上写入与读出信息的高密度存贮载体，其基本结构分为三层：基体、信息层和保护层。基体的材料可以是有机玻璃、塑料等；信息层是由极薄的金属薄膜或色素薄膜，非晶体薄膜、光磁材料等制成的；保护层是一层透明聚合物，用于防尘和防划痕。

##### （一）光盘的特点

1. 信息存贮密度极高、容量极大，是磁盘存贮密度的 50 多倍。
2. 信息的光盘存贮成本低、价格低。价格仅为同样容量磁盘的千分之一，且体积要小得多，易于保存和携带。
3. 具有随机存取特性，便于和计算机接口。光盘上的信息分布在盘层表面，读写头能迅速访问，可随机存取与快速检索，不受时间限制。
4. 可以存贮与显示多种信息。光盘既能存贮文字与数字，又能记录彩色活动图像；它不仅能用于存贮视频信息，而且还能存贮音频信息。
5. 光盘坚固耐用，存贮寿命长。光盘密封性能好，不易受周围尘土、潮气及其周围杂散电磁场的破坏，其使用寿命可达 10 年以上。
6. 便于复制。将录有信息的光盘制成凹凸模板，可以大量复制，且拷贝盘制作价格便宜，信息存贮的成本费用低。

##### （二）光盘的类型

目前已投放市场的光盘产品主要有三类：

1. 只读型光盘。只读型光盘是最早实用化的光盘。盘片在出厂前由厂家预先用激光光束蚀刻上视频、音频、数字信息，出厂后只能读上面的内容而不能添加或修改。其技术成熟、容量大、易复制、价格低，主要适用于电子出版业、联机检索系统、计算机辅助设计、办公自动化、辅助教学等。

由于只读光盘能够在微机上使用，所以得到迅速普及。而且国际上各种主要公众检索数据库都已制成只读光盘，方便人们利用。

2. 写一次型光盘。用户可以根据自己的需要自由地进行信息记录，但只能写一次，且不能修改和涂抹，所以如有变动则需要未记录部分追加记录。其技术上已经成熟，比较稳定可靠，主要应用于用户自建数据库以及文献等信息的存贮，可以作为计算机外存。

3. 可擦型光盘。这种光盘在写入信息后还可以抹掉重写新的信息，制造这种光盘难度较大。传统上有两种擦除重写方式，一是先用激光将过时的信息擦掉，再用激光束写入新信息；二是擦除和记录用两束激光分别同时进行。

可擦型光盘适用于保存更新较快的信息和信息的短期存贮。

光盘存贮技术以其众多的优点，如记录密度高、存贮容量大、工作稳定可靠、环境要求低、信息保存时间长等，吸引了人们广泛的应用，并且光盘检索系统的发展相当迅速，普及面越来越广，已成为人们生活和科研中较好的辅助工具。

## 二、全文信息存贮与检索

全文检索不仅意味着可通过计算机直接检索到全文，而且意味着可以全文中的任意符号串作为广义检索词进行全文检索运算及逻辑运算。目前的全文检索技术实际上与一般的文献检索技术十分相近，类似于一般的顺序检索，但它更早地采用了所谓“一次性扫描技术”，即文献未进行标引与索引，当提问中有若干个广义检索词时，系统不必用每一个检索词逐个地对全文进行扫描，只要将其组成一个有限自动机，然后对文献进行一次性扫描，可判定其中哪些存在于该文献中，类似于一般文献对检索词建立倒排索引的做法。在全文检索中“广义检索词”允许是单个符号或任意符号串，所以相应的倒排索引往往建立到一切具有检索可能的字符串这一级。

全文检索是以全文存贮为基础的，目前的数据库检索系统一般只提供文献的题录，如文献题名、著者名、著者通信方式、所在期刊名、卷次、页码等项，或者提供文献的文摘以供参考，用户在此基础上再根据检索线索获取原文。而全文检索一次性地解决了原始文献的直接查阅，是检索工作的一次重要变化。现在，国外已有不少数据库能够实现全文检索，但还没有达到真正意义上的全文检索，即文献中的任意字符或符号串都能成为检索词对象。全文检索技术将彻底改变人们查找文献的方式。

## 三、多媒体、超文本与超媒体

多媒体存贮检索技术是指图、文、声混合存贮与检索的技术。由于图、声，特别是运动中的图像需要大量的存贮空间，以及当图像、声音、文字同步显示时需要复杂的管理控制，所以在日常工作和生活中虽是图、文、声同步共现，但要在计算机上实现则很不容易。由于现代存贮技术与压缩存贮技术的发展，图、文、声并茂的多媒体存贮与显示已成为可能，多媒体的产品已风行世界。实际上，目前的多媒体产品大多是将图像与声音压缩后当成一个文件或一个记录存贮到计算机中，使用时即可与文本信息一样地使用，而且还借助于附加在图像或声音旁的标引信息来实现对图形和声音的检索。今后的发展则是对图形和声音的直接检索。支持多媒体数据库的应用软件，一般包括多媒体信息的采集编辑软件（加工存贮、格式转换等）、检索软件及软件输出等。

超文本就是对文本中有关内容产生联想检索的文本。超媒体则是对多媒体中部分可产生联系检索的多媒体。例如，一个系统的文本中若有“我国历史悠久、地大物博、人口众多……”，当将光标指向“我国”时，系统可立



即联想检索出“中华人民共和国及有关的其他信息”；当光标指向“人口众多”时，它即可联想检索出“中国人口的历史变迁表”。该系统就称之为超文本（能表现文本信息）或超媒体（能同时同步表现图形、文字、声音三类信息）。因此，要实现超媒体存贮，就不能完全像传统式样地按时序进行线性存贮，而必须用非线性方式把若干个可产生联想的内容组合在一起构成语义网络，以保证联想和同步“浏览”图形、文字和声音功能的实现。

#### 四、图像、视频存贮与检索

一幅图像的存贮需要大量的存贮空间。海量存贮器——光盘为图（如工程图纸、珍贵照片、图稿、指纹等）的存贮提供了存贮条件。因此，图的存贮与检索得到了发展。目前除少数系统（如指纹识别系统等）外，大多数的检索主要借助于附于图的标引信息，一幅图往往要给出6~8个关键词作为标引信息，而其检索技术与传统的文本检索并无大的区别。可见，建立图像存贮与检索系统关键在于图像的压缩存贮和复原输出技术。

视频又称之为可视图文，是指可显示在电视机与终端上的图文。视频又分为广播式视频与交互式视频。广播式视频的信息显示在电视机上，无人机交互能力；交互式视频的信息保存于计算机中，允许人机交互地传递信息，并可自由地控制信息的传递。实现交互式视频的存贮与检索，是未来存贮检索技术的发展方向。

#### 五、网络化信息检索

网络化信息检索，就是通过通信网或信息高速公路获取和享用电子信息服务的统称。在全网化通讯和信息交流的时代，不能共享网络信息资源的信息检索系统将不会有太大的发展前景，因此网络化信息检索的发展是大势所趋。

通过计算机网络系统可以连接世界上各种公用数据库和商用数据库，通过用户接口友好的信息检索工具，可以帮助用户获取希望得到的信息。

网络化信息检索系统的特点是：一般基于客户机/服务器而非基于主机，同联机系统的主机和用户的主从关系不同，客户与服务器是同等关系，只要遵循共同协议，一个服务器可被多个客户访问，一个客户可访问多个服务器。因此，网络化信息检索系统彼此可看成为客户和服务器的关系；带有图形用户界面的智能型工作站被用来取代简单终端，终端通过局域网络或广域网络而不是硬件线路与主机相连；网络的用户数量不受到限制，它是一个开放系统而不是封闭式系统，更便于增加外部信息资源。

网络化信息检索将在全球范围内实现信息资源共享，因此，全球网络将是信息存贮与检索进一步发展的方向之一。

总之，信息的存贮技术在最近的几十年里得到了迅速的发展，而且新的存贮技术也不断出现。这些存贮技术各有自己的特点和优势，可以分别在不同的领域和范围中发挥自己的特长，用户也可根据自己的需要与条件，来选择不同的存贮系统。展望未来，信息存贮技术有可能较好地将所有存贮介质融合起来，形成将磁带、磁盘、缩微品和光盘等完美结合的复合系统，从而实现各类文件图像的输入、输出、存贮检索、显示和相互转换。而这种复合系统既有极高的存贮容量，又有多样灵活的文件信息管理方式；它既能在一个用户终端上高速处理与检索存贮于多种介质上的文件图像，又能方便地把它们显示或打印出来，同时，还能通过网络与通信设备，远程传送与交换文件图像。这样的系统为各个组成部分创造了充分发挥自己特长与作用的环境。

境，因而将能够在更大范围内，更有效地实现信息的存贮和检索服务。

#### 第四节 计算机信息存贮

当前，无论是信息的印刷存贮、缩微存贮、磁存贮、光存贮还是其它新型存贮技术，都已与计算机技术紧密地联系在一起。而计算机科学本身就是一门研究信息的结构、存取、处理和应用的科学。下面对计算机信息存贮的结构及计算机存贮系统加以概述。

##### 一、计算机信息存贮的结构

###### (一) 信息的逻辑结构

信息的逻辑结构关系到计算机存取和利用信息的方式。信息在某种意义上代表着现实世界的客观现象，作为计算机处理的对象，则应抽象成能为计算机处理的数据集合。通常，数据集合中的数据元素不是孤立的，而是彼此相关的，这种彼此相互之间存在的相互关系就叫结构。人们将数据元素之间内在的、固有的联系称为逻辑结构。逻辑结构可分为两类：

1. 线性结构。这类结构的特点是：数据元素按一定顺序构成一个有限序列，除了第一个和最后一个数据元素外，每个数据元素有且仅有一个直接前趋和一个直接后继。

线性结构是一种最常用、最简单的数据结构。如单向线性表、双向循环表、环形表、数组、串、限定性的线性表、栈和队伍等，都是线性结构。

2. 非线性结构。现实世界中许多问题以非线性结构（又称离散结构）来表示要比线性结构表示明确、方便得多。非线性结构中最常用和最重要的是树型结构和网型结构。

(1) 树型结构：这类结构的特点是至少存在一个结点（元素）；除了根结点和子结点外，其他结点最多只有一个直接前趋，并有一个或不只一个直接后继。

通常，人们在研究树型结构时，又把树分为二叉树、线索树、二叉排序树、判定树、森林等来讨论。

树型结构能很好地描述信息结构的层次特性，是信息的重要组织形式之一，在计算机科学等领域应用十分广泛。

(2) 网型结构：网型结构的特点是结点间的联系是任意的，即任何一个数据元素都可以与其他元素相联结。

人们又通常分有向图和无向图来讨论网型结构。

网型结构已广泛应用于系统工程、遗传学、控制论、电子线路、统计力学及计算机网络等科技领域。

在研究线性结构、树型结构和网型结构时，一般采用下列四种方式加以记录：

(1) 序列方式：即先按段间关系规定顺序，然后按顺序排列各段的方式。这种方式实际上是以隐含的形式记录着信息的内在关系。

采用这种方式的优点是，对段进行顺序处理时效率较高。其缺点是：不易表现网型结构；对段的增加、删除或修改很不方便；寻找单个元很费时间（它适于大批处理）。

(2) 指针方式：即在段中放一指引元（又称之为“指针”）用以指引与其相关的方式。

(3) 列表方式：把数据间的关系独立出来，用行列列表表示出来。它可以用来表示各种数据结构。

(4) 混合方式：即前三种方式的混合使用，有时会达到更好的效果。例如，索引顺序就是常见的一种混合记录方式。

## (二) 信息的存贮结构

当信息由计算机进行处理时，就必须考虑其他计算机存贮器中的映象问题，这就是信息的存贮结构或物理结构。一种逻辑结构可以通过映象得到与它相应的存贮结构。信息的两大类逻辑结构——线性结构和非线性结构，分别由顺序映象和非顺序映象得到两种不同的存贮结构：顺序存贮结构和非顺序存贮结构。

1. 顺序存贮结构。在计算机中，主存贮器由字的有序序列构成，每个字包含 8 到 64 个二进制数位，可以用一个地址引用它的内容。最常用最简单的方式是以一片地址连续的存贮空间顺序存贮线性结构的数据元素，即以数据元素在存贮器中物理位置上的紧邻来表示数据元素之间的逻辑关系。在程序设计语言中，常用向量这种数据类型来描述顺序存贮结构，它利用计算地址定出数据元素的位置。

顺序存贮结构的优点是结构简单，容易实现，顺序存取速度较快，可随机存取表中任一元素。

以顺序存贮结构构造的文件，称为连续文件或顺序文件。存贮介质可以是磁带、磁盘和光盘。文件中记录的长度允许是定长的或可变长的。

2. 非顺序存贮结构。非顺序存贮结构的最大特点是不要求逻辑上相邻的元素在物理位置上也相邻，即可以用一组任意的存贮空间存贮数据元素。典型的非顺序存贮结构就是链式存贮结构。采用链式结构，在逻辑上是有序的，而在物理上则可能是无序的。

采用链式结构构造的文件，称为串联文件，存贮介质可能是磁盘、磁鼓和光盘。这种链式结构的文件建立时，无需事先确定好文件的长度，且允许文件动态增长，插入和删除记录都比较方便。

显然，对于非线性结构的信息，很难以顺序存贮结构映象；而非顺序存贮结构既能映象线性结构，又能较好地映象非线性结构。

## 二、计算机信息存贮

### (一) 计算机存贮系统的结构

计算机是存贮与处理信息的机器。计算机存贮系统主要由存贮器及存贮管理软件构成。

1. 存贮器。存贮器是组成计算机系统的重要部件，输入的信息必须存贮在某一规定的位置上，一旦需要时才能很方便地使用。因此，需设存贮器。

存贮器可以分为内存贮器（主存贮器）和外存贮器（辅助存贮器）两部分。

内存贮器存贮容量较小，但存取速度快；外存贮器的存贮容量大，但存取速度较慢。人们通常是把常用的数据放在快速存取的主存贮器里，供随时调用；而把不常用的数据存入容量大、速度慢的外存贮器里。

目前所用的内存贮器是磁芯存贮器。这种存贮器往往由数千万字节的存贮单位组成，为了识别这些存贮单位，每一字节应编好地址，数据则按地址存贮在规定的位置上，或从规定的位置取出。存入和取出所需时间，称为存取时间。

由于磁芯存贮器容量有限，无法容纳帐目信息之类的大量数据，因此这类数据应存贮在费用低廉而容量很大的外存贮器内。外存不和计算机的操作直接发生关系，只有在必要时才移到磁存贮器进行处理。

外存贮器主要有两种：磁带存贮器和随机存取存贮器。

磁带存贮器是一种顺序存贮器，它是按照磁带录音的原理进行存贮的。当存取数据时，必须从头按顺序检索到所需位置，故其存取时间因存贮位置不同而相差很大。为了克服这一缺点，就研制出了随机存取存贮器。这种存贮器是根据地址进行查找的，所以无论存贮位置如何，其存取时间大致相等。

随机存取存贮器的载体是磁鼓和磁盘。磁盘是一种快速的大容量外存，是大型软件系统必不可少的硬件基础。磁盘部分可更换的磁盘存贮器称为磁盘组件，现已研制出在一台驱动器上可安装九个磁盘组件的存贮器，大大提高了存贮容量。

为提高存取速度，人们又在中央处理器和内存贮器之间增加了一个小容量高速的缓冲存贮器，用以存放中央存贮器当前最常用的指令和数据，与主存交换信息。

2. 存贮管理。存贮器是计算机系统的关键性资源，能否合理有效地使用它，取决于计算机操作系统中的存贮管理模块和文件管理模块。

不同的操作系统可能采取不同的存贮管理方式，但其目的都是一致的，即为了解决存贮空间的分配，扩充主存贮器容量，实现信息的存贮、修改、检索、共享和保护，提高存贮空间的利用率。

在操作系统中，通常讨论的存贮管理内容是对内存贮器的管理，而文件管理的内容是对外存贮器的管理。

(1) 内存的管理：对内存的分配通常有两种分配策略：静态分配策略和动态分配策略，分别采取静态重定位和动态重定位技术。我们把作业在装入过程中进行地址变换（即把逻辑地址变换成内存中的物理地址）的方式称为静态重定位；而把作业在执行进程中，当访问到某一指令或数据时，才进行地址变换的方式称为动态重定位。具体的存贮管理方案有：单一连续区分配、分区、分页、分段、段页式分配等。不同的存贮管理方案有不同的分配方式、保护信息的措施和共享信息的可能性。

(2) 外存的管理：对外存的管理，是由操作系统中的文件管理系统——存取和管理信息的模块，通过提供为用户建立文件，对文件存贮空间进行组织、分配、保护，负责文件的读取、转贮、控制等功能来实现的。用户只需知道他们的文件名和操作命令，便可按名存取和操作，而不必关心文件究竟存放在什么物理位置上。

对外存空间的分配，可根据信息的逻辑结构采用两种分配策略，一是分配连续的存贮区域，二是分配不连续的物理块。

(3) 虚拟存贮器：对于内存贮器容量的扩充，通常借助于自动覆盖和交换技术来达到。一个大作业提交给系统，其中一部分信息在内存，另一部分信息暂时留在外存，需要时若内存空间已满，系统将自动依据某种算法选择暂时不用的信息从内存移出，再将需要处理的信息调入内存。此时呈现给用户的是一级存贮器的映象，这就是所谓的虚拟存贮器的概念。有了虚拟存贮器后，正在运行的多道作业的全部地址空间总和可以大于主存的容量。

要实现虚拟存贮器，要有动态重定位地址变换机构，一定容量的内、外存贮器和相应的管理软件。

## 二、文件信息的存贮方式

文件是在逻辑上具有完整意义的信息集合。不同的信息集合提交给计算机系统进行处理，通常是以文件为单位并以文件名来识别的一个作业或一个任务。文件的逻辑结构可分为两种：一是记录式文件，这类文件由若干逻辑记录组成，逻辑记录是在逻辑上具有独立含义的一个信息单位；二是流式文件，这类文件是一个顺序的信息流集合。文件的存贮结构密切地依赖于存贮设备的物理特性，但对用户而言，文件概念的引入，使得用户可以用统一的方式去存取不同存贮介质上的信息。

一个计算机系统必须提供一定的外存空间，供用户存放以文件形式组织的信息。一般说来，一个以程序设计语言编写的文件作为一个作业，从文件输入到信息输出，须经过编辑、编译、链接和运行四个不同的作业步，在外存以三种形式的用户库来存贮。

### （一）源程序库

以源语言形式出现的计算机程序库称源程序库。任何经编辑作业步处理的信息都可以存放在源程序库中，而无须每次上机重新建立文件。由源程序库提交系统执行，必须经过编译、链接和运行三个作业步。

### （二）相对形式库

在具有一定规模的、支持多道程序设计的计算机系统中，用户可以将经过编译作业步处理而产生的相对模块存入在磁盘或磁带上建立的相对形式库。

用户可以将所有已调试好的子模块存入相对形式库，需要经常改动的主模块存入源程序库；或者将已调试好的主模块和其它子模块存入相对形式库，而新添加的未调试好的子模块存入源程序库。这样存贮的优点是，提交系统执行时，节省了时间较长的编译作业步，从而节省了整个作业的执行时间。

### （三）执行形式库

用户可以将经编辑、编译和链接三个作业步的目标模块存入在磁盘或磁带上建立的执行形式库。当用户的程序比较成熟，则无须改动，而每次运行需要更换不同的数据集合时，最好采用这种方式，可有效地缩短机时，加快执行速度。

总之，计算机存贮是计算机技术发展的基础，也是其它存贮技术赖以发展的条件。无论是信息的缩微存贮、磁存贮、光盘存贮，还是更新型的存贮及检索，都是建立在计算机技术的基础之上，与计算机技术紧密结合，才能寻求更大的发展空间。可见，计算机存贮及技术乃是计算机科学、信息科学长久研究的课题。

## 第八章 计算机信息检索

计算机信息检索就是利用计算机对信息进行存贮与检索。相对于手工检索,计算机检索无论在检索途径、检索速度,还是在检索范围、检索时差方面,均有很大的优势,因而得到了广泛的应用。目前,运用计算机检索系统进行文献信息检索已越来越普遍,而且,随着科学技术及计算机检索技术的进一步发展,计算机信息检索将逐步取代手工检索,而成为文献信息检索的主要手段。

### 第一节 计算机信息检索概述

#### 一、计算机信息检索的发展阶段

计算机信息检索的发展过程是与计算机技术及其它现代科学技术的发展过程紧密相关的。计算机用于信息检索始于 50 年代初,在 40 多年的发展历史中,计算机信息检索大体经历了三个发展阶段。

##### (一) 脱机检索阶段

此阶段是从 50 年代中期到 60 年代中期。自 1946 年 2 月世界上第一台电子计算机问世以来,人们一直设想利用计算机查找文献。进入 50 年代后,在计算机应用领域“穿孔卡片”和“穿孔纸带”数据录入技术及设备相继出现,以它们作为存贮文摘、检索词和查询提问式的媒介,使得计算机开始在文献检索领域中得到了应用。

1954 年,美国海军兵器中心首先采用 IBM-701 型计算机建立了世界上第一个科技文献检索系统,实现了单元词组配检索,检索逻辑只采用“逻辑与”,检索结果只是文献号,1958 年,美国通用电器公司将其加以改进,输出结果增加了题名、作者和文献摘要等项目。1964 年,美国化学文摘服务社建立了文献处理自动化系统,使编制文摘的大部分工作实现了计算机化,以后又实现了计算机检索。同年,美国国立医学图书馆建立了计算机数据库,即医学文献分析与检索系统,不仅可以进行逻辑“或”、“与”、“非”等种运算,而且还可以从多种途径检索文献。

这一阶段主要以脱机检索的方式开展检索服务,其特点是不对一个检索提问立即作出回答,而是集中大批提问后进行处理,且进行处理的时间较长,人机不能对话,因此,检索效率往往不够理想。但是,脱机检索中的定题服务对于科技人员却非常有用,定题服务能根据用户的要求,先把用户的提问登记入档,存入计算机中形成一个提问档,每当新的数据进入数据库时,就对这批数据进行处理,将符合用户提问的最新文献提交给用户,可使用户随时了解课题的进展情况。

##### (二) 联机检索阶段

此阶段是从 60 年代中期到 70 年代初。由于计算机分时技术的发展,通信技术的改进,以及计算机网络的初步形成和检索软件包的建立,用户可以通过检索终端设备与检索系统中心计算机进行人机对话,从而实现对远距离之外的数据库进行检索的目的,即实现了联机信息检索。

这个时期,由于计算机处理功能的加强,数据存贮容量的扩大和磁盘机的应用,为建立大型的文献数据库创造了条件。例如美国的 DIALOG 系统(DIALOG 对话系统)、ORBIT 系统(书目情报分析联机检索系统)、BRS 系

统（存贮和信息检索系统）、欧洲的 ESA-IRS 系统（欧洲航天局信息检索系统）等都是在此时期开始研制并逐步发展起来的，并且均在国内或组织范围内得到实际应用。

可以说，联机检索是科技信息工作、计算机、通讯技术三结合的产物，它标志着 70 年代计算机检索的水平。

### （三）网络化联机检索阶段

此阶段是从 70 年代初到现在。由于电话网、电传网、公共数据通信网都可为情报检索传输数据。特别是卫星通信技术的应用，使通信网络更加现代化，也使信息检索系统更加国际化，信息用户可借助国际通讯网络直接与检索系统联机，从而实现不受地域限制的国际联机信息检索。尤其是世界各大检索系统纷纷进入各种通信网络，每个系统的计算机成为网络上的节点，每个节点联接多个检索终端，各节点之间以通信线路彼此相连，网络上的任何一个终端都可联机检索所有数据库的数据。这种联机信息系统网络的实现，使人们可以在很短的时间内查遍世界各国的信息资料，使信息资源共享成为可能。

可以说，联机网络和检索终端几乎遍及世界所有国家和地区，使得国际联机信息检索的发展达到了相当高的水平，开展商业性国际联机检索服务的大机构已达 200 余家，像美国的 DIALOG 信息公司已成为全世界最为著名的联机检索服务机构。

计算机信息检索的实现，大大方便和加速了信息资源的交流和利用，并对社会经济的发展和人们的科研方式产生了深刻的影响，从而也极大地促进了科技的进步。

## 二、我国计算机信息检索的发展

我国计算机信息检索的研究是以 1974 年开始的。1975 年，由中国科技情报所和北京图书馆联合发起，组织了 500 余单位，花了三年时间编制成收录有 10 余万条主题词的综合型《汉语主题词表》，为发展汉字信息检索打下了物质基础。1976 年利用引进的美国《政府研究报告》磁带进行计算机检索试验获得成功，接着机械工业部情报所、北京文献服务处等许多单位先后利用国外文献磁带开展定题检索和国内联机信息检索服务。1980 年由中国建筑技术发展中心文献部等十一个单位联合在香港租用国际联机检索系统终端，为国内各有关部门的科研、设计施工、生产教学和管理进行国际联机检索服务。1981 年，北方科技信息研究所在北京建立了第一个国际联机终端用户。1982 年中国科技情报研究所建立了中国药文学文摘汉字数据库。1983 年中国科技情报研究所设立国际联机检索数据终端，与美国 DIALOG 系统和欧洲的 ESA-IRS 系统连接，开展国际联机检索服务。1983 年，“科技情报计算机检索系统”的开发任务列入“六五”国家重点工程项目，从此我国计算机信息检索事业进入有计划发展阶段。

“七五”期间，我国科技信息计算机检索系统，作为国家优先发展的计算机应用系统之一，得到了迅速的发展。中央各部委、直辖市、大部分省、中国科学院和高等院校系统的科技信息机构，已配备大中小型计算机 110 多台，建立了具有建库、检索、刊物编辑等配套功能的信息系统 30 多个。我国已引进国外文献数据库约 70 种，三分之一已提供或具备条件提供联机检索服务，文献总量超过 2000 万篇。我国已有近 50 个城市开通国际联机检索业务，与国外 12 个大型信息检索系统联机，能够检索 500 多个数据库。我国自建数

据库约 400 种，包括中文文献数据库 30 个和品种多样的事实型、数据型数据库。

目前，微型机在我国已相当普及，微型机在信息工作和人们生活中得到越来越广泛的应用，利用微机进行信息检索正在迅速地开展。

## 第二节 计算机信息检索系统

计算机信息检索系统就是利用计算机的有效存贮和查找能力来进行信息的分析、组织、存贮和查找的系统。

### 一、计算机信息检索系统的构成

计算机信息检索系统主要由三个部分构成，即硬件部分、软件部分和信息数据库。

#### （一）硬件部分

硬件是指以计算机主机为中心的一系列机器设备，包括主机、外围设备以及与数据处理或数据传送有关的其他设备。

主机可以是大型、中型计算机，也可以是微型机，它负责整个系统的运行和管理，一般应具有较高的运算速度和较快的响应速度，同时还要有较强的逻辑运算能力及对大量文献或数据的处理能力。

外围设备包括外部存贮器、输入输出设备等。存贮器用以存贮原始数据、程序和计算结果。外部存贮器主要包括磁带、磁盘、光盘等。输入输出设备包括各种必要的的数据录入装置和输出装置，输入设备有数据录入设备、键盘、鼠标器、光笔、缩微扫描器等；输出设备有显示终端、打印机等。

其它与数据处理或数据传送有关的硬件设备还有运算器、控制器等。运算器主要用以对数据进行运算、比较、判断、组合等加工处理，从而完成计算机的计算工作，其运算结果传送到存贮器中。控制器主要控制输入输出设备、内外存贮器和运算器，使之按照程序的规定，产生所需要的命令，并控制命令的执行和信息处理。

广义地讲，硬件设备还应包括电源设备和通信设备及网络设备。

#### （二）软件部分

计算机软件又称计算机程序，是指控制计算机进行各种作业的一系列指令和进行“人机对话”及各种数据的存贮和传输的“翻译”规则。软件都存放在磁盘或磁带上，也可固定存放在主存贮器中，一旦需要就可取出并送到控制器中，控制器则按程序中的指令操纵各个部件进行工作。软件包括操作系统软件、语言编译软件、应用软件和用户软件等。没有这些软件，计算机就不能运行。

#### （三）数据库

数据库是计算机信息检索的重要组成部分。它是以二进制代码形式在计算机存贮设备上（如磁带、磁盘和光盘等）合理存放的相关数据的信息集合，通常由存贮信息记录及其索引的若干文档组成。不同的数据库，存贮不同主题、数量、时间和类型信息的。

目前，大多数计算机信息检索系统都存有四种类型的数据库：文献型数据库、数值型数据库、词典型数据库和全文型数据库。

1. 文献型数据库，又叫书目型数据库。记录文档中主要存入的是原始文献的书目，例如：原始文献的篇名、作者、文献出处、文摘、关键词等，主



要用于查询各种文献资料的书目性线索、文摘等。

2. 数值型数据库。数据库的记录存入各种调查数据或统计数据。它是一种以自然数值形式表示、计算机可读的数据集合，这些数据是从文献中分析、概括、提取出来，或以调研、观测及统计工作中直接获得的数据。数值型数据库主要用于查询各种有关的数字、参数、公式等。

3. 词典型数值库。这类数据库的文献记录主要介绍一些有关公司、团体或名人的情况，供用户查询某一事物发生的时间、地点、过程或简要情况，或诸如化学物质名称、结构、俗称和化学物质登记号之类指南性信息，故又称之为指南型数据库。

4. 全文型数据库。这类数据库的文献记录存入了原始信息的正文。通过它可以直接检索出原始信息的全文，而不必进行二次检索，从而大大方便了用户，也提高了信息的有效利用。但是，目前大多数的全文型数据库的记录中还没有能包含原始文献中的图表。同时，全文型数据库需要占用更多的存储空间，因而，目前的国际联机信息检索系统中，多半还是一些涉及金融、商业、市场或时市等方面信息的全文型数据库，而涉及科技方面的全文型数据库还不多见。

## 二、计算机信息检索系统的服务方式

目前，计算机信息检索系统主要有以下几种服务方式：

### （一）追溯检索服务

追溯检索是根据用户的检索要求，对检索系统中积累多年文献资料的数据库进行检索，查找一定时间范围以内或特定时间以前的文献，这是联机检索中最常用的检索方式。

运用追溯检索，一次就能了解一定时期以来积累的全部有关信息，非常适合于科技研究工作中对信息的需求。

### （二）定题信息服务

这种检索是将用户预先确定好的检索提问式长期保留在检索系统的中央计算机中，中央计算机在相应数据库更新时，自动按保留的检索提问的要求将存贮到数据库中有关的最新信息记录检索出来提供给用户。这对长期以来某一专题研究或开发，而又需要长期跟踪其最新动态的用户特别有用。

### （三）光盘服务

随着光盘技术的发展，一些计算机检索系统，尤其是大型国际联机检索系统还开展了光盘服务。例如，从1987年开始，DIALOG系统已经可以向用户提供部分数据库的光盘盘片，并可提供与之配用的软件和所需设备，盘片可以每季节更新一次。这种局部系统可以独立进行检索。如果不能完全满足信息需求时，用户可以随即通过联机进行补充检索。

## 第三节 国际联机信息检索

国际联机检索是用户利用检索系统网络的终端设备，通过国际通讯网络与国际联机检索系统的中央计算机，从检索系统的数据库中查找所需的文献信息资料。

### 一、国际联机检索系统的组成

国际联机检索系统由计算机中心、数据库文档、检查终端、计算机通信网络几个部分组成。

### （一）计算机中心与数据库

安装在联机检索中心的计算机都是大型的分时装置，可供几十个甚至几百个终端同时进行信息存取，其功能是完成信息的存贮、处理和检索等操作，对整个系统的运行进行管理和控制。供检索用的数据库文档存贮在磁盘上，磁盘驱动器通过控制装置与计算机相连，可以随时检索存贮在磁盘上的信息内容。

计算机中心通过通信控制器、调制解调器与通讯网络和终端相连。

### （二）检索终端

终端是用户与主机进行人机对话的工具，是检索人员将用户的检索要求等信息通过一系列的检索指令传递给远端的计算机，并从计算机接受信息的一种设备。联机检索终端包括终端机、电话机、调制解调器、显示器、打印机等装置。

到目前为止，我国设置的联机检索终端按照连接的数据库系统不同，可分为国际联机检索终端和国内联机检索终端两类。国际联机检索终端是直接通过国际通信网和卫星与国际大型联机检索系统相连的；国内联机终端是通过国内专用电话线路与国内自建的数据库检索系统连接的。按照通信线路类型来分，终端又可分为用户电报终端和公用数据终端两种。前者使用用户电报线路，比较容易解决问题，但传输速度低；后者使用数据或电话通信线路，速度快，可靠性强，费用低。我国多采用微型机作为联机检索终端。

### （三）计算机通信网络

计算机通信网络一般由计算机、终端设备、数据通信网（包括调制解调器和通信线路）、通信控制器（是计算机与远程终端或网络中其它计算机进行通信的控制设备）组成。它可以将地理位置分散的资料实现实时的集中管理，并达到数据库资源、硬件资源、软件资源的共享目的，从而提高检索系统的利用率。计算机通信网络在联机信息检索系统中占有重要的地位，只有有了计算机网络，才可能出现信息检索的联机系统。

在通信网络的线路方面，我国终端用户主要使用用户电报网线路和公共电话交换网线路。前者的优点是可利用本单位的用户电报线路和电传机作终端，缺点是费用较高，速度慢。后者的优势是传输速度快，质量好，国内大多数国际联机检索终端都采用用户电话网线路，当用户多、数据传输量大时，宜采用专线方式；当用户少时、数据传输量小时，宜采用电话拨号方式。这两种方式各有利弊，应视具体情况而定。此外，还有第三种线路，即公共数据网线路，它与上述两种线路不同，具有持续时间短、误码率低、传输速度快的优点。我国的联机检索系统的通信线路正在向公共数据网的方向过渡。

## 二、国际联机信息检索的优点

国际联机信息检索克服了地理位置的限制，又兼有一般计算机检索的特长，使其具有许多优点。主要表现在以下几个方面。

### （一）检索范围广

建立在卫星及其它现代通信网络基础之上的国际联机检索系统，使得用户打破了使用信息的空间界限，检索者不论在哪一国家或地区，只要有一台联机终端，就可检索网上所有有机检数据库中的各类信息。现在许多联机检索系统通过国际或国内通信网络，与设置在世界各地的终端相连，供用户检索，而速度和效果不受用户地理条件的影响。可以说，国际联机检索系统最大限度地体现着人类信息资源共享的现代特点。

## （二）数据库种类多、数量大

许多国际联机检索系统都拥有各种学科领域的数据库，系统的数据库种类多，信息存贮量大。有的国际联机系统甚至拥有几百个至上千个数据库，收录几千万篇或上亿篇文献或记录，可满足各种专业和行业的用户需求。例如，仅 DIA-LOG 系统就有近 300 个数据库可供检索，信息的类型可以是文献型、数值型或者事实型，文献的出版形式可以是期刊、图书、报告、专刊、会议录等，它们以世界上所有重要文种出版，来自世界上所有国家地区。据统计，世界上公开发行的文献，80% 都可以通过国际联机系统查到。

## （三）检索速度快

由于计算机具有高速处理信息的能力，所以它可以使用户在最短的时间里，以最快的速度获得所需文献信息。国际联机检索的主机，大都能每秒执行几千万条指令，一般用户发出命令 5 秒钟内即可将命中文献的篇数显示到用户的终端屏幕上，检索一个课题只需几分钟。

检索速度快还体现在另一方面，即从原始文献的出版到它们在检索系统中得到报道的时差比较短。各数据库的更新周期一般都不长，多则一个季度，少则半月乃至一周，因此最新文献很快就能在数据库中得到反映。当前，绝大多数书本式检索工具都是由计算机在其机检数据库的基础上编制的。显然，书本式检索刊物报道文献要比机检数据库晚一些。在我国，往往已收到报道某一篇文章的磁带版数据库，还未收到登载该篇文章的原版期刊。如果利用国际终端，进行联机检索，还会得到更新的文献线索。

## （四）检索途径多，检索效果好

联机检索系统可以提供众多的检索途径，用户可根据需要运用多种检索方法进行查检。用户能够自行确定检索词，或用布尔逻辑算符和位置算符将多个检索词进行组配查找，也可指定任意信息属性，如主题词、著者、出版单位，文献篇名、期刊刊名、出版年代等为检索途径。而且可以通过人机对话，扩大或缩小检索范围，不断修改与调整检索策略，以达到满意的查全率和查准率，确保检索的最佳效果。

## （五）具有灵活的服务方式和输出方式

一般大型的联机检索系统都提供回溯检索、定题检索、联机订购原文乃至联机通信等多种服务项目。如可通过联机系统订购原始文献的复印件。因为用户通过联机检索后，获得的都是二次文献，当用户需要文献资料的原文而在当地馆藏中又找不到时，则可利用联机检索系统利用原文的服务，向提供原文的机构订购原文的复印件或缩微平片。另外，联机检索的结果可联机显示并打印，也可脱机打印，还可将联机结果存盘，以便重新编辑整理使用。输出格式有多种选择，既可输出数据库中的完整文献记录，也可输出篇名、出处等某些字段，或根据需要安排安段的排序，并编成专题目录、信息资料通报等。有的系统还可输出缩微形式的原始文献。

## （六）检索全面

由于当今各学科的资料分布异常分散，因而给全面查找某学科的资料信息带来了一定的困难。据统计，一般只有 50% 左右的资料刊登在相应学科的杂志上，而其余部分则分散在其他学科的杂志上，对边缘学科、新型学科更是如此。而联机检索可以很容易地做到在检索系统的不同数据库中进行搜索，只需几分钟到十几分钟就可以得到全面而具体的结果，大大节约了查寻时间，保证了查全率。

### 三、世界著名的国际联机检索系统简介

#### (一) DIALOG 系统

DIALOG 系统又称 DIALOG 对话系统，因研究人机对话检索方式而得名。它是目前世界上最大的国际联机检索系统，也是我国信息用户使用最多的系统。1963 年系统创建初期，曾是美国洛克希德导弹和空间公司下属的一个情报科学实验室。当时规模很小，实验室仅有一台 IBM360/30 型计算机，用于进行信息检索试验。1966 年，DIALOG 系统与美国国家宇航局的艾姆斯研究中心合作进行联机检索的实验，通过电话线路，从一台终端上检索研究中心图书馆存贮 20 多万条文献记录的数据库。1972 年，DIALOG 系统开始成为国际联机检索系统，正式对外开展联机文献检索服务业务。之后，其规模不断扩大，影响越来越广，联机检索业务逐渐发展到加拿大、欧洲、日本、澳大利亚乃至第三世界国家。

1981 年 6 月，该系统正式宣布作为洛克希德公司的子公司，独立经营检索业务。其总部设在美国加利福尼亚州的帕洛阿尔托 (Palo Alto) 市。目前，该系统的联机服务网遍布世界 70 多个国家和地区的 200 多个城市，共设有联机检索终端 25000 余台。系统中各类型数据库数量的发展速度相当快，到 1987 年 6 月，系统总计拥有 300 多个数据库，存贮的文献超过一亿条，占世界机检文献总量的 50% 以上。

DIALOG 系统各文档涉及的学科范围相当广泛，包括自然科学、应用科学、社会和人文科学及各种综合学科和交叉学科等。另外还包括关于市场行情、时事报道、数据统计、经济预测和机构名录等方面的信息。文献类型有：书刊、报纸、学位论文、研究报告、专利、标准等。其服务方式除定题服务、追溯检索等常规服务外，还提供多文档检索服务、家庭检索服务、电子邮件服务、原文订购服务、光盘检索服务等。

1988 年，洛克希德公司将 DIALOG 系统以总共 3.53 亿美元转卖给一家报纸和电视通讯公司 (Knight-Ridder 公司)。

#### (二) ORBIT 系统

ORBIT 系统即书目情报分时联机检索系统，是美国系统发展公司 (SDC) 与美国国际部于 1963 年共同开发的联机检索系统。总部原设在美国加利福尼亚州的 Santa Monica 市，是仅次于 DIALOG 系统的第二大国际联机检索系统。

ORBIT 系统于 1965 年在美国国内实现联机检索，1974 年发展为国际联机检索系统。1987 年脱离美国系统发展公司，与英国的 Infoline 系统共同归属于英国 Pergamon 公司集团新建的 Pergamon Orbit Service 公司。新公司成立 orbit 检索服务部，并将总部移至弗吉尼亚州的 Reston 市。

目前该系统已拥有 2 万多台终端，遍及世界许多国家和地区的大城市。该系统拥有 90 多个数据库，分两种类型：书目型数据库和指南型数据库，涉及的专业范围有化学、能源、工程、生物工艺学、计算机科学、商业经济以及其他一些自然科学和社会科学。存贮的文献类型有：图书、期刊、会议论文、专利、年度报告、政府法规、产品规格、统计数据、报纸评论等。其特点是对汽车工程、石油、化工、医学、环境科学、安全科学、运动科学、专利等方面的文献收录较齐全，并对一批使用价值较高的数据库拥有独家经营服务权。

#### (三) BRS 系统

BRS 系统即存贮与信息检索系统，成立于 1976 年，1979 年投入商业性运

营，总部设在美国纽约。

BRS 系统 80 年代末已拥有 150 个公用数据库，40 多个私人数据库，其中大部分为书目数据库。数据库内容包括社会科学、人文科学、商业经济、教育、医学、生物化学、工程技术等。文献类型不仅包括了各类书目文献，还收集有关专刊、政府报告、标准、工业产品说明等。该系统在 1980 年率先实现了全文检索，建立了 20 多个全文数据库，对当天出版的期刊也可立即存入。在期刊数据库中，有 100 多种可实现联机全文检索。目前，文献存贮量达 5000 多万篇，用户 4 万多个。该系统还以优惠政策为我国用户服务，即系统中所有数据都对我国开放，并按美国国内最低标准收费。同时，为满足我国用户需要，增加了五个有关产品和标准的数据库，并可用缩微方式提供原始资料复制品。

#### (四) ESA-IRS 系统

ESA-IRS 系统是欧洲航天局信息检索服务处的英文缩写。简称 ESA 系统，是欧洲最大的文献检索系统，也是世界上最重要的检索系统之一。该系统是欧洲航天局的一个下属单位，成立于 1964 年，当时叫做欧洲空间文献中心 (SDS)，总部设在巴黎。1973 年改名为 ESA-IRS，并将总部迁往意大利罗马附近的弗拉斯拉蒂 (Frascati) 市。除西欧外，系统的联机检索终端还遍及美洲、北非、中东、澳大利亚和亚洲的许多国家和地区。

ESA 系统有 100 多个文档，总计存贮 3000 万条书目数据。专业范围包括航空、航天和宇宙空间、天文学、天体物理、环境与污染、地理学、海洋学、化工、经济、管理和社会科学等多种学科，但没有专利文献数据库。ESA 系统包括的数据库有近一半与 DIALOG 系统或 ORBIT 系统系统相重叠。但某些数据库独具特色，如法国文摘通报数据库、原材料价格数据库等。它在装库方面也具有其他系统所没有的特色，即把规模大的数据库，不管是最新部分，还是追溯部分，都维持在一个文档中，避免了用户分别检索几个子库的麻烦。

ESA 系统提供的服务除回溯检索、定题检索和联机订购原文等项目外，还提供用户之间的快速通信业务，并可联机结算用户帐目。

#### (五) STN 系统

STN 系统即国际科技信息网络系统，是由美国化学文摘社 (CAS)、德国卡尔斯鲁厄能源、物理、数学专业情报中心 (FIZ) 和日本国际化学情报协会 (JAICI) 三家联合经营的国际联机检索系统，1983 年底开始筹建，1986 年初正式对外服务。其服务中心设在德国的卡尔斯鲁厄、美国的哥伦布和日本的东京三处。每一个中心分工建立一部分数据库，同一个数据库只存贮在一处。美国的化学文摘社主要建立化学文摘数据库；德国能源、物理、数学专业情报中心主要建立能源、物理、数学等方面的数据库；日本化学情报协会主要把现有的化学化工方面的数据库转入 STN 系统。

到 1989 年，三个中心共向系统投入 83 个数据库，约 6000 万条记录。虽 STN 系统所收录的数据库与其它系统有很多重复，但 STN 系统还收录了一些独特的数据库。如由德国专利局提供的德国专利数据库，它是目录、全文、图像三结合的数据库。另外，STN 系统拥有世界上第一批联机图像数据库，如美国化学文摘社的化学物质结构图形库、贝尔斯登有机化学大全结构数据库等。

STN 被设计成为全球范围可供检索的系统，每天工作 24 小时，其不同国家合作经营的方式使它在国际联机检索服务市场中具有较强的竞争力。

