

The background of the cover features a close-up of a hand holding a pen, poised to write on a document. The document has some text, including 'Clarification', 'Club Membership', 'College', 'Costs', and 'IMPORTANT'. Below the document, a computer keyboard is visible, with the keys illuminated by a warm, golden light. The overall color palette is dominated by warm tones of orange, yellow, and brown.

# 信息检索 教程

许增福 吴春玉 金秋颖 主编

石油工业出版社

## 《信息检索教程》编委会

主 编：许增福 吴春玉 金秋颖

副主编：梁文义 金元宪 李瑞滨

吴点顺 冯志鹏

## 前 言

当今，世界正经历着一场前所未有的变革，工业时代正迈向信息时代。以现代传感和测量系统、现代通信系统、高性能计算机和智能系统以及现代控制和显示系统为代表的智力工具的出现，正把近代工业社会生产力转变为现代信息社会生产力，社会生产将提高到空前的水平。

处在这一迅猛发展的时代，无论是科技人员还是普通的老百姓都必须不断更新自己的知识，必须不断采用新的智力成果以变革生产方式，必须及时掌握科技、社会的发展信息以保证研究工作的先进性。同时，也借助于高新技术使自己的生活和环境质量不断提高。

因此，人们就需要从浩如烟海的信息海洋中及时地获取符合要求的、准确的、全面的信息。随着计算机的普及、计算机网络技术的应用，以及人们对信息的需求，信息检索技术得到了普遍的重视。信息检索是一门实用技术，是每一个生活在信息社会的人们都应掌握的技能。

近几十年来信息检索技术有了长足的发展，从手工检索到计算机检索，以至今天的网络检索。本书全面介绍了信息检索基础理论和基本概念、国内外的主要文献检索工具及特种文献的检索；详细地阐述了计算机信息检索的基本知识、信息检索技术、检索策略；还介绍了国际互联网络和光盘检索、文献的利用等方面的内容。适合作普通信息检索用户及高校检索课教材使用。

随着信息技术的发展，检索手段在不断更新。当本书完成时，一些数据库检索系统可能又有了新的版本，尤其是 Internet 上的检索工具更是名目繁多，花样迭出，因此提醒读者在使用本书的过程中，注意浏览和借鉴其他资料，以求全面系统地掌握这

一领域的知识。

本书的第一、三、四、五章由吴春玉编写；第二章由梁文义和吴点顺编写；第六、七、八章由金秋颖编写；第九章由李瑞滨编写；第十章由金元宪和冯志鹏编写。全书的整体构思、结构设计、修改和定稿由许增福完成。

本书在构思、组织、整理及写作过程中，得到了孙桂媛、韩颖、王继龙、佟喜峰等同志的多方帮助，仅在此表示衷心的感谢。

本书虽经著者努力，但错误、缺点在所难免，希望广大读者批评指正。

编 者  
2001年5月

# 目 录

<b>第一章 信息检索基础知识</b> .....	( 1 )
第一节 信息检索概述.....	( 1 )
第二节 信息资源.....	( 12 )
第三节 信息检索工具.....	( 17 )
第四节 信息检索语言.....	( 22 )
第五节 信息检索程序、策略及效果评价.....	( 37 )
<b>第二章 国内检索工具</b> .....	( 42 )
第一节 国内检索工具概述.....	( 42 )
第二节 国内检索工具体系.....	( 43 )
<b>第三章 国外综合性检索工具</b> .....	( 47 )
第一节 美国《工程索引》.....	( 47 )
第二节 美国《科学引文索引》.....	( 52 )
第三节 前苏联《文摘杂志》.....	( 56 )
第四节 日本《科学技术文献速报》.....	( 60 )
<b>第四章 国外专业性检索工具</b> .....	( 64 )
第一节 美国《化学文摘》.....	( 64 )
第二节 英国《科学文摘》.....	( 70 )
第三节 美国《石油文摘》.....	( 75 )
<b>第五章 特种文献及其检索</b> .....	( 85 )
第一节 专利文献及其检索.....	( 85 )
第二节 会议文献及其检索.....	( 92 )
第三节 科技报告及其检索.....	( 94 )
第四节 标准文献及其检索.....	( 98 )
<b>第六章 计算机信息检索概述</b> .....	( 100 )
第一节 计算机信息检索发展过程.....	( 100 )
第二节 计算机信息检索技术.....	( 101 )

第二节	电子信息资源	(103)
第四节	数据库类型	(107)
第五节	计算机检索过程与基本功能	(114)
第七章	国际联机信息检索系统	(120)
第一节	国际联机检索概述	(120)
第二节	DIALOG 联机检索系统	(125)
第八章	光盘信息检索	(135)
第一节	光盘检索概述	(135)
第二节	KR OnDisc 检索系统与 Ei Compendex * Plus 光盘数据库	(137)
第三节	中国学术期刊全文数据库	(148)
第四节	《中文科技期刊光盘数据库》检索	(152)
第九章	Internet 信息检索	(156)
第一节	Web 简介	(156)
第二节	国外著名搜索引擎的使用方法	(160)
第三节	中文搜索引擎的使用方法	(165)
第十章	文献的利用	(168)
第一节	科学研究与文献的利用	(168)
第二节	文献资料的积累与整理	(169)
第三节	科技论文的撰写	(171)
附录		(176)
参考文献		(181)

# 第一章 信息检索基础知识

## 第一节 信息检索概述

### 一、信息检索概述

随着人类的进步，社会的发展，科学技术有了突飞猛进的发展，同时也产生了大量的信息。尤其是人类迈进 21 世纪，有人说这个世纪是信息时代。人们生存在信息的海洋中，任何人的衣食住行都与信息息息相关，在这种环境中如何能提高生存质量？如何能跟上时代的步伐，推动社会进步？其中的关键问题之一是怎样捕捉信息、利用信息，因此生活在这一时代的人们都应该掌握一种技能即信息检索技术。

#### 1. 信息

##### (1) 信息的含义

我们每天都同信息打交道，经常使用信息这一词汇。可是，目前学术界还没有统一的定义。信息的英文词汇是 Information，对这一词的解释有几十种，不同的领域有不同的解释。诸如：“信息就是信息，既不是物质也不是能量，而是与物质和能量相并列的三大要素之一。”“信息是使人增加知识，并能向人们解释事物的客观存在。”“信息是消息，是信号，是数据，是情报，是知识。”等等。在国内信息领域比较认同的看法是信息是事物存在的方式和运动状态的表征，信息并非事物本身，而是事物发出的体现它存在和运动状态的信号和消息。信息普遍存在于自然界、社会各界和思维界。因此，信息可以理解为通过信号传来的消息。信息是被传递的数据、事实、符号的集合，客观事物存在反映的物质质量。

信息有传递性、存储性、可识别性，并消除不确定性、替代性和可共享性。

## (2) 信息的种类

信息从不同角度划分，划分出很多种类。

1) 按信息的广义内涵划分，可分为自然信息和社会信息。

自然信息是指宇宙间，自然界客观存在或随机发生的各种生命信息、动植物界信息、物质物理信息等。

社会信息是指人类和社会维系生存、生产和发展过程中所产生、传递和利用的信息。这一类信息就是人们平常所说的信息。

2) 按信息的社会属性划分，可分为政治信息、军事信息、经济信息、科技信息、管理信息、生活信息等。

3) 按信息的传递范围划分，可分为公开信息、半公开信息和非公开信息。

4) 按信息的加工程度划分，可分为一次信息、二次信息、三次信息。

5) 按价值观念划分，可分为有价值信息和无价值信息。

6) 按运动状态划分，可分为动态信息和静态信息。

动态信息是指像新闻、气象等具有较强的实效性、随时都在变化着的信息。而静态信息是指历史的、不在变化的信息，比如：各类文献、历史资料等。

7) 按载体形式划分，可分为文献信息、声像信息、多媒体信息。

文献信息是指将文字记录在各种载体上的信息。声像信息是指各种声音和图像的信息，如：广播、电视及卫星图片等。多媒体信息是近几年发展起来的一种新兴的信息传播形式，它将文字、声音、图像融为一体，提供给人们更加完善的信息。

## (3) 相关概念

1) 情报。

情报在《辞海》中解释为：①以侦探手段或其他方法获得的有关敌人军事、政治、经济等各方面的情况，以及对这些情况进



行分析研究的成果，是军事行动的重要依据之一。②泛指一切最新的情况报道。在人们的印象中，情报一词带有军事行为，具有保密性的含义。随着时代的进步，情报这一词已无法适应时代的发展。1992年国家科委决定将科技情报改为科技信息，许多情报所已改为信息所，这样更具有时代的特点。

#### 2) 知识。

知识是人们在改造客观世界的实践中所获得的认识和经验的总结。

知识来源于实践，通过实践，客观事物必须反映到人的大脑中，反映的深入，知识的深化，从现象到本质，从感性到理性，从而掌握了事物的规律性，也就产生了知识。信息是认识事物的媒介，生产知识的原料。所以，知识生产离不开对信息的接收。

#### 3) 文献。

随着科学技术的发展，出现了各种不同的载体材料，发明了各种各样记录知识信息的方式，文献大量涌现，文献概念的外延也在不断扩大。国际标准化组织《文献情报术语国际标准》(ISO/DIS52217)对文献的定义是：“在存储、检索、利用或传递记录信息的过程中，可作为一个单元处理的、在载体内、载体或依附载体而存储的信息或数据的载体。”这不仅包括了传统的书刊、文稿等，而且包括了缩微品、音像资料、机读资料和电子出版物等新型的信息知识载体。所以可以将文献理解为：记录有信息或知识的一切载体。

#### 4) 资料。

为工作、学习、生产和科学研究等参考需要而收集或编写的一切公开或内部的材料。

#### 5) 资讯。

资讯是海外学者对 Information 的译名，与信息有相同的含义。

### 2. 信息检索

### (1) 信息检索的含义

信息检索是指将信息按一定的方式进行加工、整理、组织并存储起来，再根据信息用户的需要找出有关信息的过程。它的全过程又叫信息存储与检索。这是广义的信息检索的含义，主要是对信息工作者而言的。狭义的信息检索则仅指后半部分，即用户根据需要，借助检索工具，从信息集合中找出所需要的信息的过程。

### (2) 信息检索的意义

信息检索是查找信息的方法和手段，它能使人们在浩如烟海的信息海洋中迅速地、准确地、全面地查找所需的信息。可以说信息检索对人们的学习、生活和工作等方面都有非常大的作用。

#### 1) 信息检索在科学研究中的作用。

人类的科学技术发展与进步都是在前人已有的知识和成果之上产生的，正如著名科学家牛顿所说的那样：“如果我看到的，我所发现的要比笛卡尔和培根远大一点的话，那就是因为我站在巨人的肩膀上的缘故。”科学研究的每一项成果都是在总结归纳了大量信息的基础上完成的，因此信息检索对科技人员来说是十分重要的。具体表现在以下几方面：

①充分地开发和利用信息资源。

②节省科研人员的时间，提高效率，减少人力物力等方面的费用。

③防止自身知识老化，再学习的工具。

④继承和借鉴前人的研究成果，避免重复研究和少走弯路。

#### 2) 信息检索在生活中的作用。

有人认为：“我既不搞科研，也不搞管理，信息检索可能与我无关。”在过去可能是这样，但由于网络的出现，改变了我们的生活。现在有些发达国家和一些地区已实现了衣食住行等诸多问题都可以在信息网络上解决，因此离开了信息检索将很难生活。

## 二、信息检索原理

### 1. 信息检索原理

人类的信息检索行为总是从特定的信息需求而开始，并在特定环境和信息检索系统中完成，这里所说的环境包括产生需求的环境、信息检索系统的运行环境和其他制约因素。特定的检索系统包括完成检索过程所需的一定设施和工具，它可以是图书馆、信息中心或信息经济人，也可以是某种工具书（如：文摘索引、目录、资料集、手册、词典等）或机读信息源（如：各种机读数据库）。

人类的信息需求千差万别，获取信息的方法也各种各样，但信息检索的基本原理却是相同的，可以把它最本质的部分概括为一句话：对信息集合与需求集合的匹配与选择（见图1-1）。

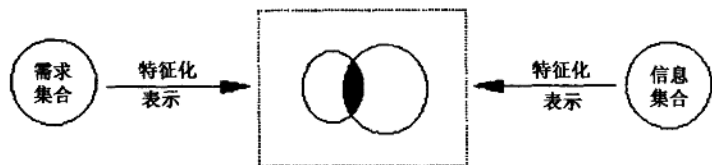


图1-1 信息检索基本原理示意图

### 2. 信息检索手段

根据信息检索的基本原理，实现信息检索的基本方式可分为传统信息检索和现代信息检索，传统信息检索，简称“手检”；现代信息检索，简称“机检”。

#### (1) 传统信息检索

即手检，是检索人员利用手工检索工具手翻、眼看、大脑思维判别、索取原始文献的一种方式，其优点是：①检索条件简单，成本低；②在检索过程中可以随时获取反馈信息，及时调整检索策略；③可对不同的检索工具同时进行对比，从而提高检索质量；④可以参阅检索工具中的附图。其缺点是：①速度慢、效

率低，检出的文献款目必须抄录；②手工检索工具提供的检索点有限，很难进行多元检索；③难于找到涉及几个概念组合的多主题的文献。手检的检索方法有以下两种：

### 1) 直接检索。

这是人们常用的检索方法，即利用一次文献进行检索。人们常常在需要信息资料时，直接到图书馆查阅各种图书、期刊及其他资料，从中找到所需的信息。

### 2) 间接检索。

就是利用各种检索工具（目录、题录、文摘）查找信息的方法。这是文献检索最常用的方法。

## (2) 现代信息检索

即机检，是检索人员利用计算机检索系统查找文献的一种检索方式。所谓计算机检索系统包括数据库技术、计算机技术和通讯技术等。机检可以克服手检的缺点，但机检对设备条件的要求比较高，所需的投资比较大。计算机检索已从单机检索、联机检索发展到今天的网络检索，并正向着智能化的方向发展。

### 1) 单机检索。

单机检索是计算机检索的最初形式，随着计算机存储介质的发展变化，也在发生不断的变化。目前主要有三种形式：计算机磁盘检索、计算机磁带检索和计算机光盘检索。

### 2) 网络检索。

网络检索是近十几年发展起来的，它目前主要有三种形式：图书馆局域网的检索、联机检索及 Internet 网上信息检索。

## (3) 传统信息检索与现代信息检索的比较

现代信息检索是在传统信息检索的基础上发展起来的，因此两者的检索原理是相同的，但在 Internet 网上的信息检索技术与以往的计算机检索技术又有很大的差别，主要有以下几方面的异同点：

### 1) 检索语言方面。

检索语言就是在检索信息时所使用的专门的语言词汇，手工

检索与计算机检索所使用的检索语言基本相同，不同的是手工检索使用规范语言较多，而计算机检索则可以使用较多的自然语言，Internet 网上检索一般使用自然语言。

#### 2) 概念的组配方面。

手工检索无检索算符，检索时各种概念之间用人脑进行组配，用计算机检索时则是利用各种逻辑算符进行概念的组配，Internet 网上检索概念组配方式还不成熟，目前一般只用布尔逻辑算符进行组配。其中计算机检索的概念组配最为严谨，手工检索的概念组配最灵活。

#### 3) 检索入口方面。

也就是手工检索工具中的检索途径，是利用根据信息的某种特征所编制的索引进行的检索，如：题名、著者等。手工检索的检索入口较少，而计算机检索和 Internet 检索的检索入口较多而且非常灵活，计算机检索过程中一个信息的所有特征都能做为检索入口，如：年代、语种、文献类型、各种代号等，检索时还可以多途径同时进行检索。

#### 4) 检索结果方面。

三种检索相比，手工检索的检索结果查准率最高，误检率最低，而查全率较低。计算机检索的结果查准率要低于手工检索，误检率要高于手工检索，查全率则较高。Internet 网上检索结果误检率较高，查全率也较高。

### 三、信息检索的研究范围与对象

信息检索作为一门学科，有它自己的研究范围和对象，也有自己的理论、方法和技术。从总体上看，信息检索的研究对象是比较明确具体的，研究范围广泛而边界有些模糊，理论和方法已经逐渐形成体系。

#### 1. 信息检索的研究范围

信息检索的研究范围包括一切与信息存贮与检索有关的系统、过程、理论和方法。一切可供存贮和检索利用的信息类型，如文献、数据、事实、知识、声音、图形等；各种信息检索系统

及其运行过程，如信息采集、标引、组织、存贮、处理、匹配、传送等各种过程中使用的方法；以及在信息检索实践和研究的基础上形成的各种理论和假设，均包括在这个范围内。信息类型侧重于文献，其次是数据和事实。

## 2. 信息检索的研究对象

主要有以下几方面：

1) 信息检索理论。它主要包括检索语言与标引理论、信息检索的数学模型、知识表示理论、相关性理论，以及有关的哲学问题。

2) 信息检索系统。它是实现信息检索的物质基础，是现实的研究对象，主要研究它的结构、功能、演变，它的设计开发技术、管理维护技术和评价技术，还研究它与其他信息系统乃至整个外部世界的关系。其中，数据库是信息检索系统的核心部分之一。数据库的建造和维护是一类非常重要的信息技术。数据库的质量直接影响信息检索系统的功能和效率。

3) 计算机信息检索。它涉及到许多计算机设备、软件技术、存贮技术革新、检索技术、系统管理和经营知识、市场营销技术等，是一门综合性很强的技术。

4) 检索策略与方法。它是用户从信息检索系统中获取有关信息所必需的。好的检索策略是检索成功的必要前提，计算机的应用为检索策略和方法的发展提供了有力的支持。近 30 年来，涌现了许多新颖而有效的检索技术和方法，如布尔检索、位置检索、截词检索、加权检索、聚类检索等，人们利用、研究、评价和完善现有的各种检索策略和方法，研究开发新的更有效的策略和方法。

5) 用户研究与培训。用户是信息检索系统的生命，无论是系统的研制开发、管理维护、功能和服务的扩展，还是系统评价，都离不开用户研究工作。用户培训是用户研究的继续，是与用户建立紧密联系和发展新用户的一种非常有效的措施。

此外，还有自动标引、自动分类和自动摘录，以及相关设备

等研究领域。

#### 四、信息检索系统

信息检索系统是由一定的设备和信息集合构成，面向一定的用户，具有信息采集、组织、存贮、选择和传播等功能。

按着不同的标准，信息检索系统可以划分为不同的类型。

##### 1. 按设备划分

根据信息存贮与检索使用的设备，可分为以下类型：

1) 书本式检索系统。它包括一切以书刊形式提供，为人们查找各种信息或数据的出版物，如文摘杂志、题录或索引刊物、参考工具书等。

2) 卡片式检索系统。它包括一切以普通卡片存贮和查找信息的工具，如图书馆内的各种卡片目录、管理部门的各种卡片档案。与书本式系统相比，它较便于信息的累积和更新，更适于单位或个人自建自用，成本较低，在信息检索和资料管理中曾发挥过重要作用。目前，它已受到计算机目录系统的有力挑战，面临着被取代的命运。

3) 计算机信息检索系统。早期的计算机信息检索系统以机器穿孔卡片为存贮介质。后来，发展到以磁带为外存贮器，利用计算机顺序读取磁带数据库中的数据，故又称顺排档检索系统。20世纪60年代中期以后，随着磁盘和大容量存贮器的投入使用，大量数据可以随机存放在磁盘上，基于磁盘的检索系统正式出现，它可以支持回溯检索，人们又称之为倒排档检索系统。计算机信息检索系统进一步发展为联机检索系统和光盘检索系统，联机检索系统通常由大型计算机系统、数据库和一批检索终端等设备构成。检索终端通过通讯设备与系统的主机连接，并在中央处理机控制下工作，查询系统的数据库。

光盘检索系统出现于20世纪80年代，以大容量的光盘存贮器为数据库的存贮介质，利用光盘驱动器读取光盘上的数据。

此外，还有20世纪90年代发展起来的Internet信息查询系统。

## 2. 按功能划分

许多信息系统都含有检索的机制和过程。它们存贮的数据文件或信息库，在响应用户提问时被处理，然后生成和输出答案。在这个意义上我们可以把它们看成是广义的检索系统。

1) 文献检索系统，它是狭义的信息检索系统。目前绝大多数信息检索系统都属于此类型。由于文献信息数量巨大，品种复杂繁多，连贯性强，故文献检索系统中的信息存贮通常采用结构化的文件结构。检索的目的是要获得书目引文和原始出版物，而不是现实的数据。大型商业性的文献检索系统通常都是专门开发的，检索功能较强，即适用于书目数据和文本信息的存贮与检索，也可以用于事实和数值型数据的存取。一方面检索系统主要面向文献资源的管理者和利用者，其用户主要是科研和教学人员。此外，它还能对各种管理和决策活动提供有效的信息支持。

2) 数据库管理系统 (Data Base Management System, 简称 DBMS) 是为任何数据库的建立、操作和维护而设计的计算机软件。它最显著的特征是：存贮的数据具有确定的结构，即把数据或信息存贮在结构化的数据文件中，系统只能使用那些结构完全确定的数据项。这是它与文献检索系统的主要区别。DBMS 主要面向数据查询和基层事务管理，其主要用户是行政职员或事务管理员。

3) 自动问答系统 (Question Answering System, 简称 QAS) 是一种能直接回答人们提出的具体问题的一种计算机系统，曾被称为事实检索系统，目前已归入专家系统的范畴，问答系统的主要用途是支持或代替专家决策，为知识的采集、存贮和利用服务。它采用的知识库技术和自然语言处理技术对文献检索系统和 DBMS 有潜在的应用价值。问答系统是信息检索系统的高级形式。

4) 管理信息系统 (Management Information System, 简称 MIS) 是建立在数学方法 (统计理论、计算方法)、管理科学和信息技术的基础上，对各种用于管理的信息进行采集、加工、分



析、整理、保存的一种人一机系统，它以各有关基层业务部门的信息系统和事务处理系统为基础，把来自各子系统的数据进行加工分析，最后向管理者提供有关计划、管理、作业控制、预测等方面的信息，一般以文件报表为主。20世纪70年代以后，随着计算机与通讯技术的发展，MIS逐渐进入一体化阶段，并迅速应用于社会各个领域，它虽然保留有信息存贮与检索功能，但更主要的是一种管理决策工具，面向管理人员而不是一般用户。数据的统计分析已成为它的主要功能。

5) 决策支持系统 (Decision Support System, 简称 DSS) 是管理信息系统的高级形式，主要面向中、高层管理和执行决策，尤其是支持半结构或非结构化决策，为决策者提供一种分析问题、构造决策模型和模拟决策过程及其效果的决策环境。DSS的主要功能特征是：①采集、存贮、编辑和检索综合性信息或文献；②便利最终用户洞察非结构化问题和作出判断；③利用统计模型、经济模型和会计模型去分析综合信息，并为最终用户生成观点鲜明的状态报告和预测；④利用专门的管理科学模型去评价各种决策方案，并形成推荐性的策略。

### 3. 按应用领域划分

信息检索系统或具有查询检索功能的信息系统几乎在各个领域都有应用，且可能以不同形式出现。

科技信息与图书馆管理领域中有图书信息检索系统、全文检索系统、图书馆自动化系统等。科学研究与工程设计领域中有科研管理系统、实验数据与解析系统、分子信息检索系统等。行政管理领域有政务信息系统、办公自动化系统等。经营管理领域有生产管理信息系统、商品流通管理系统、物质管理系统、专利管理系统等。财政金融方面有银行信息系统、财务管理系统、电子资金转账系统等。

医疗卫生领域有医院管理信息系统、计算机辅助诊断系统、病人监护系统、药品配方管理系统等。

教育领域有计算机辅助教学系统和教务管理系统。