高等学校研究生教学用书

金秋颖　韩　颖　王园春　◎编
王文广　◎审

数字信息检索技术

SHUZI XINXI
JIANSUO JISHU

石油工业出版社

高等学校研究生教学用书

数字信息检索技术

金秋颖　韩　颖　王园春　编
王文广　审

石油工业出版社

内 容 提 要

本教材针对研究生学习以研究为主的特点，重点放在学术性数字信息检索技术上。本书共分五章，主要介绍了信息资源及其检索基础，常用中、英文数据库系统，专利与标准文献检索及论文写作方法等内容。

本书可作为高等院校研究生和高年级本科生检索课教学用书，也可作为教师、科研人员从事教学、科研信息检索的学习指导书。

图书在版编目(CIP)数据

数字信息检索技术/金秋颖，韩颖，王园春编.
北京：石油工业出版社，2006.2
(高等学校研究生教学用书)
ISBN 7-5021-5419-1

Ⅰ.数…
Ⅱ.①金… ②韩… ③王…
Ⅲ.计算机应用-情报检索-研究生-教学参考资料
Ⅳ.G252.7

中国版本图书馆 CIP 数据核字(2006)第 005949 号

出版发行：石油工业出版社
(北京安定门外安华里 2 区 1 号 100011)
网 址：www.petropub.cn
总 机：(010)64262233 发行部：(010)64210392
经 销：全国新华书店
排 版：北京乘设伟业科技排版中心
印 刷：石油工业出版社印刷厂

2006 年 2 月第 1 版 2006 年 2 月第 1 次印刷
787×1092 毫米 开本：1/16 印张：8.5
字数：213 千字 印数：1—2000 册

定价：18.00 元
(如出现印装质量问题，我社发行部负责调换)

前　言

目前，随着网上搜索引擎、数字图书馆、大型数据库的不断建设以及网络查询技术的日益普及，已经打造了一个数字信息检索的新平台。与传统的文献检索形式相比，数字信息检索的优越性表现在两个方面：第一，这是一个跨地区、无国界的信息空间系统，数字化的文字、图片、声像文件无论存储在什么位置，都可以通过 Internet 互相连接，使用户可以超越时空限制，突破“图书馆”的约束，在任何地方、任何时候都能获取这些信息。第二，Internet 数字信息检索以智能为基础，以全文检索为特征，大大突破了传统检索的限制。因此，检索和利用数字信息已成为教师、科研人员、学生迫切需要掌握的技能之一。

高等院校数字信息检索课程的建设与发展，对培养研究生的自学能力、获取信息的能力、创造能力和动手能力，具有积极的作用。该课程教学具有专业性、实践性等基本特点。所谓专业性，是指各学校结合本校的系科专业，讲授带有明显专业色彩的信息检索和利用的途径与方法；所谓实践性，是指这门课程重于实际的资讯查找，重于方法和技能的训练，强化课程中的实践检索环节，把理论知识联系到操作应用，落实到课题检索的实际过程。

本教材针对研究生以学习研究为主的特点，重点放在学术性数字信息检索技术上，有针对性地编写了五章。第一章介绍信息检索概念、检索语言、检索系统、计算机检索技术、检索程序及策略、检索效果评价方法。第二章介绍中国学术期刊网、万方数据资源系统、高效财经数据库系统、电子图书等常用中文检索系统。第三章着重介绍 EBSCO、Sringer Link、EI、SPE、PA 外文检索系统。第四章重点介绍专利与标准文献的检索。第五章介绍信息的整理与利用、论文写作方法等内容。

本书第一章由金秋颖和韩颖编写，第二章由金秋颖编写，第三章由金秋颖、王园春编写，第四章由韩颖、王园春编写，第五章由韩颖编写。全书由金秋颖、韩颖负责内容的策划和汇总统稿。该书得到了王文广教授的关心和指导，并由王文广教授承担主审工作。

由于 Internet 网处于动态发展状况，书中介绍的网址可能有的发生了变动，读者如果发现个别网站无法登录，请用大型搜索引擎等方法查询。由于作者水平有限，书中难免存在错误及疏漏之处，恳请广大读者和同行不吝指正。

编　者

2006 年 1 月

目　录

第一章　信息资源利用基础 …… (1)
第一节　信息资源概述 …… (1)
第二节　信息检索 …… (7)
第三节　检索语言 …… (9)
第四节　信息检索系统 …… (11)
第五节　数据库及其种类 …… (12)
第六节　计算机检索技术 …… (14)
第七节　检索程序、策略及效果评价方法 …… (18)
复习思考题 …… (20)
第二章　常用中文检索系统 …… (21)
第一节　CNKI 系列数据库 …… (21)
第二节　中文科技期刊数据库 …… (30)
第三节　万方数据资源系统 …… (33)
第四节　INFOBANK 高校财经数据库系统 …… (38)
第五节　电子图书 …… (43)
复习思考题 …… (47)
第三章　国外常用检索系统 …… (48)
第一节　EI《工程索引》 …… (48)
第二节　Springer Link 检索系统 …… (53)
第三节　EBSCO 数据库检索系统 …… (57)
第四节　SPE 论文索引检索系统 …… (63)
第五节　Petroleum Abstracts(石油文摘)检索系统 …… (65)
复习思考题 …… (67)
第四章　特种文献检索系统 …… (68)
第一节　专利文献检索 …… (68)
第二节　标准文献及其检索 …… (83)
第三节　搜索引擎 …… (90)
复习思考题 …… (94)
第五章　文献信息的利用 …… (95)
第一节　文献信息的收集、整理、分析与利用 …… (95)
第二节　科技综述与科技述评的写作方法 …… (101)
第三节　学术论文的写作方法 …… (103)

第四节　开题报告与结题报告的写作方法 …… (111)
第五节　毕业论文的写作方法 …… (115)
第六节　学术论文投稿及相关信息的获取 …… (121)
复习思考题 …… (126)
参考文献 …… (127)

第一章　信息资源利用基础

现代信息技术迅猛发展，信息已成为人类社会发展的一种驱动力，人们越来越重视对信息资源的有效开发与利用。信息是一种极其重要的社会财富，信息同物质、能量构成了人类社会的3大重要战略资源。物质提供材料，能量提供动力，信息提供知识与智慧。因此，信息已成为促进科技、社会、经济发展的新型资源，它不仅有助于人们不断地揭示客观世界，深化人们对客观世界的科学认识，消除人们在认识上的某种不定性，而且还源源不断地向人类提供生产知识的原料。

第一节　信息资源概述

一、基本概念

1. 信息

信息作为一个科学术语最早出现于通信领域，20世纪中叶后被引入哲学、信息论、系统论、控制论、情报学、经济学、管理学、计算机等领域。不同学科的学者、专家以及有关领域对信息的定义都是从信息的受体、内涵和控制论等角度对信息的属性进行的描述。因此，我们说信息是事物属性的再现。信息不是事物本身，而是由事物发出的、体现它存在和运动状态的信号、消息、指令和数据等所包含的内容。即广义的信息可定义为“信息是事物属性的表征”；而狭义的信息则可定义为“信息是指系统传输和处理的对象”。

2. 知识

知识是人们在改造世界的实践中所获得的认识和经验的总和，知识的本质则是认知活动中的主体与客体的动态关系。知识是人类在认识和改造世界的社会实践中获得的对事物本质认识的成果和结晶，是人的主观世界对于客观世界的概括和如实反映，是人类通过信息对自然界、人类社会以及思维方式与运动规律的认识，并通过人的大脑进行思维重新整合使信息系统化而构成知识。因此，人类不仅要通过信息感知世界、认识世界和改造世界，而且还要根据所获得的信息组成知识。由此可见，知识是信息的一部分。

人类现有的知识可分为4大类，即：

(1)知道是什么的知识(Know－what)，是指关于事实方面的知识，这类知识通常被近似地称为信息。

(2)知道为什么的知识(Know－why)，是指自然原理和规律方面的科学理论，这类知识的生产往往是由专门的研究机构形成。

(3)知道怎么做的知识(Know－how)，是指技艺或能力方面的知识，被称为技术诀窍或专有技术。许多企业的技术情报和商业秘密被归入这类信息。

(4)知道是谁的知识(Know－who)，是指谁知道和谁知道如何做某些事的信息，这在社会高度分工的经济时代中显得尤为重要。这类知识比任何其他种类的知识都更隐藏在企业内部。

3. 文献

文献是记录一切人类知识信息的载体。文献由4要素构成,即信息、载体、符号系统和记录方式。四位一体不可分割,缺少任何一个都不能构成文献。

由信息、知识和文献三者的概念可知,三者之间的关系是密切相关的。信息是物质存在的方式、形式和运动规律的表征。人脑对事物属性的感知形成信息,人们对信息集合加工、整理形成系统化表现,形成人类社会实践的知识,知识被记录在载体上形成文献,文献被人类广泛传播,运用在理论和实践中,又产生新的信息、知识和文献。如此循环不断创新形成各种各样的新知识,从而推动人类社会前进。由此可见,信息、知识、文献在社会系统中表现出一种不间断的延续性。

4. 信息资源

在人类社会和自然界的运动发展过程中,产生着各种各样的信息。每一天都会有各种形式的信息层出不穷,这种大量的、客观存在的、人们直接或间接开发利用的信息集合总称为信息资源。

5. 数字信息资源

数字信息资源是信息资源的一种,是数字化了的信息资源,即以数字的形式,把文字、图形、图像、声音等多种形式的信息存放在光、磁等非印刷型介质上,以数字信号的形式传输,并通过相应的计算机和其他外部设备再现出来的一种信息资源。数字信息资源是电子信息资源的主体。本书中讲到的电子信息资源检索也主要是指对数字信息资源的检索。

6. 信息技术

支持人们识别、接收、传递、加工、处理、利用信息的技术称之为信息技术。传感技术、通信技术、计算机技术和控制技术是信息技术的4大基本技术,其中计算机技术和通信技术是信息技术的两大支柱。

7. 信息素质

信息素质(Information Literacy)一词最早是由美国信息产业协会主席 Paul Zurkowski 在1974年给美国政府的报告中提出来的。他认为,信息素质是人们在工作中运用信息、学习信息技术、利用信息解决问题的能力。

美国图书馆协会认为:信息素质是人们知道什么时候需要信息并找到、评价及有效地利用所需信息的能力。信息素质较强的人知道如何学习,因为他们了解知识是怎样组织的,知道如何找到信息,他们能够终生学习,因为他们能够发现所有与自己职责相关的决策所需要的信息。信息素质的内涵具体包括能意识到准确和完整的信息是决策的基础;了解信息需求及问题所在;制定信息检索策略;掌握信息检索技术;能评价信息;能根据实际用途组织信息;将新信息融会到现有知识结构中。

21世纪是网络信息和知识大发展的世纪,在信息化社会中,无论是个人还是企业,信息素质是谋生存、求发展的重要因素。对于现代研究人才来讲,只有具备信息素质才懂得在信息化社会中如何去获取、加工、存储、检索和利用信息,使其拥有不断学习和持续发展的能力。

二、信息资源的构成

信息资源的构成可以从不同的层面和角度来划分。

1. 按照信息的出版或加工形式划分

信息出版类型一般是指记录有知识的文献出版类型。一般将出版物文献划分为图书、报

纸、期刊、会议文献、专利文献、科技报告、学位论文、技术档案、产品资料、标准文献和政府出版物。

(1)科技图书。

科技图书大多是对已发表的科研成果、生产技术和经验或者基本知识领域系统的论述或概括,它往往以期刊论文、会议论文、研究报告及其他第一手资料为基本素材,经过作者的分析、归纳、组织而编写成的。不少科技图书的内容还包含一些从未发表过的研究成果或资料。

科技图书的特点是:内容比较系统、全面、成熟、可靠,具有一定的新颖性;但编辑出版时间过长,传递信息的速度太慢,包含的内容一般只是反映2~5年以前的研究水平。

科技图书是综合、积累和传递科技知识,教育和培养科技人才的一种重要工具,它可以帮助人们比较全面、系统地了解特定领域的历史和现状,可以将人们正确地领入自己所不熟悉的领域,还可以作为一种经常性的查考工具。从信息检索角度来看,科技图书一般不作为主要检索对象。研究人员利用图书的比重比较小。美国有的信息专家曾经对美国各大学的科学家和英国电气工程师们进行过调查,发现在他们所阅读的各种科技文献中,图书的比重分别占19%和14%。

(2)期刊。

期刊是一种以印刷形式或其他形式逐次刊行的,通常用数字或年月顺序编号,并准备无限期地连续出版下去的出版物(ISO 3297—1986)。

广义的期刊则包括一切定期刊行或不定期刊行的连续性出版物,如杂志、报纸、年度报告、年鉴、丛书以及学会的会议录、学报和纪要等。

科技期刊在科学技术活动中一直起着非常重要的作用,是科学交流的主要工具。科技期刊具有以下特点:数量大、品种多、内容丰富多样;出版周期短,报道速度较快;发行、流通广泛,连续性强,伴随着相应学科领域的发展而发展。

(3)会议文献。

会议文献是指在国内外各种学术会议上交流的论文,以及由此汇编成册内部交流或公开出版的文献。

会议文献的主要特点是:传递信息比较及时,传递的信息针对性较强,它反映了某学科、专业的最新成果和发展现状及趋势,是研究工作不可缺少的情报源。

(4)专利文献。

专利是用法律来保护科学技术发明创造的制度。专利文献是专利制度的产物,是一切与专利制度有关的各种专利文件的统称,其中包括发明说明书、专利说明书、专利局公报、专利文摘、专利分类与检索工具书以及申请专利时提交的各种文件(如请求书、权利要求书、有关证书等)与专利有关的法律文件和诉讼资料等。狭义的专利文献一般是指专利局颁布出版的各种发明说明书或专利说明书及其所派生的各种二次文献。

专利文献的特点是:数量巨大,覆盖面广;格式统一,措辞严谨;描述对象具体、单一;技术内容新颖、可靠;文件类型多,重复量大,是重要的技术经济信息源。

(5)科技报告。

科技报告是研究或设计单位向提供经费的上级部门提供的关于某项研究或设计任务完成情况及财务消耗情况的总结报告。

科技报告的特点:从形式上看,科技报告的出版形式比较特殊,每份报告自成一册,篇幅长短不等,有连续编号,装订简单,出版发行不规则。从内容上看,科技报告的内容比较新颖、详

尽、专深。

(6)学位论文。

学位论文是高等院校或研究机构培养的学生为获得某种学位而撰写的科学论文,一般有学士论文、硕士论文和博士论文。学位论文中除了少数可能发表在期刊或其他出版物以外,多数是不出版的。每篇学位论文有一复本保存在授予学位的学校图书馆,可供查阅。

(7)技术档案。

技术档案是指在生产建设中和科技部门的技术活动中形成的,有一定的工程对象的技术文件的总称。其内容包括:任务书、协议书、技术经济指标和审批文件,研究计划、方案、大纲和技术措施,有关的技术调查材料(原始记录及分析报告)、设计计算、试验项目和方案、数据和报告、设计图纸、工艺卡片以及应入档文件。

(8)产品资料。

产品资料是指国内外各厂商为推销产品而印发的商业宣传品,包括产品样本、产品目录、产品说明书、厂商介绍、厂刊或外贸刊物、技术座谈资料等。

(9)标准文献。

标准文献是以文件形式出现的标准化工作成果。经过公认的权威当局批准的标准化工作成果,可以采用文件形式或规定基本单位(物理常数)这两种形式固化下来的文件。标准化是为了有关各方的利益,特别是为了达到最佳的经济效果,并适当考虑到使用条件和安全要求,在有关各方的协作下,进行有步骤的特定活动所制定并实施各项规则的过程。

标准文献的特点是:制定、审批有一定的程序;适用范围非常明确专一;编排格式、叙述方法严谨统一,措辞准确;技术上具有较充分的可靠性和现实性;对有关各方有约束性,在一定条件下具有某种法律效率;规定明确的有效时间,需要随着技术的发展而不断修订、补充或废除。

(10)政府出版物。

政府出版物是各国政府部门及其所属机构所发表的文件。它的内容广泛,概括起来可分为行政性文件和科技文献两大类。行政文件包括国会记录、司法资料、方针政策、规章制度、决议、指示以及调查统计资料等;科技文献包括各部门的研究报告、技术政策文件等。

2. 按照信息加工层次划分

人们在利用、传递信息过程中,为了及时报道和揭示信息,对信息进行了不同层次的加工。按加工程度可将信息分为一次文献、二次文献和三次文献。

(1)一次文献,即以作者本人的生产与科研工作成果为依据而撰写的,并已公开发行进入社会流通使用的原始文献,如专著、学术论文、科技报告、会议论文、专利文献、学位论文等。一次文献的特点是具有学术上的新观点、新发明、新技术、新成果,提供了新的知识信息,是创造性劳动的结晶,有直接参考、借鉴和使用的价值,是人们检索和利用的主要目标。

(2)二次文献,即将大量、分散、无序的一次文献收集起来,按照一定的方法进行整理、浓缩和加工,使之系统化而形成各种目录、索引和文摘,编制成具有多种检索途径的检索工具。因此,二次文献仅是对一次文献进行系统化的压缩,无新的知识产生,具有汇集性、检索性的特点。它的重要性在于提供了检索一次文献的线索。因此,二次文献又称为检索性文献。

(3)三次文献,即根据一定的目的和需求,在大量利用一次、二次文献的基础上,对有关知识进行综合、分析、提炼、重组而再生的信息资源,如词典、手册、百科全书、年鉴、各种教科书及综述等。因此,三次文献具有综合性高、针对性强、系统性好、知识信息面广的特点。三次文献又称为参考性文献,有较高的使用价值,可直接参考、借鉴和利用。三次文献源于一次文献,又

高于一次文献，是一种再创性文献。

从文献的角度看，一次文献是人们检索与利用的主要对象，二次文献是文献信息的检索工具，三次文献是人们考查数据、事实信息的主要信息源。

3. 按照信息内容划分

(1)文献信息源，即存储语言文字形式信息的各种载体的集合。文献型信息源是目前信息内容最丰富、最可靠的信息，是人们使用最多的信息源。

(2)非文献信息源，包括数值型信息源、声像型信息源、多媒体信息源和实物及口头信息源。

① 数值型信息源：存储数据形式信息载体的集合。

② 声像型信息源：存储声音或图像信息载体的集合，如磁带、广播、电视。

③ 多媒体信息源：是一种时代发展的产物，它集文字、声音、图像于一体，以光盘或 Internet 网上资源的形式出现，是目前发展最快、数量最多的一种信息源。

④ 实物及口头信息源：实物信息源是指自然实物和人工实物中所含信息的集合，口头信息源是指在交流、讨论、报告过程中所含的信息集合。

4. 按照信息的载体和传输形式划分

(1)按载体形式划分。

① 印刷型：以纸张为介质，通过铅印、油印、胶印、复印等手段记录信息的载体。

② 缩微型：以感光材料为介质，通过缩微照相手段记录信息的载体。

③ 机读型：以磁性材料为载体，通过编码和程序设计，由计算机输入和输出的信息。

④ 声像型：以电磁材料为载体，借助特殊设备，直接将声音和图像等信息记录下来的一种动态信息。

(2)按照信息传输形式划分。

① 网络信息源：各种网络上的信息集合。

② 非网络信息源：不用通信设施就能获得的信息集合。

三、数字信息资源

1. 信息资源的类型

我们知道，数字信息资源是通过电子计算机等以数字信号来传递的信息资源。数字信息资源是电子信息资源的主体。我们目前所说的电子信息资源主要指的是数字信息资源。数字信息资源按照不同的标准可划分为若干类型。

(1)按照信息的载体划分。

① 联机网络信息资源：20 世纪 60 至 70 年代，世界上发达国家和地区相继建立起计算机联机信息服务系统，如美国的 Dialog、欧洲共同体的 ESA 和德国的 STN 系统等，都为全世界联机用户提供了丰富的数字信息资源。但长期以来，由于费用昂贵，许多用户不敢问津。90 年代以来 Internet 迅猛发展，网络资源十分丰富，价格相对低廉，且有许多免费资源。这使得人们越来越多地从网络上检索、获取信息，通过网络来共享全球的数字信息资源。

② 单独发行的信息资源：以光盘出版物为主。

(2)按照信息的媒体形式划分。

① 文本信息资源：普通的文本信息资源的知识单元按线性顺序排列。阅读时，人们跟随文本的线性流向吸收其中的养分，遇到不懂的地方或想要知道详细情况时，就得暂时中断阅

读，去查阅有关参考资料。这样就打乱了文本固有的线性配置格局，在读者的头脑中形成了相互参阅的知识单元网状结构。然而，用户不易掌握和追踪这种网状结构，更难以对其修改和补充，仅靠手动、眼看、心记是具有极大的局限性和片面性的，超文本的出现为解决这一问题提供了手段。

② 超文本信息资源：超文本是指一种人—机交互的友好系统。用户利用计算机可以增删内容，用户的想法可随时存入数据库中，也可随时检索、调用。超文本是一种通信，它可以组织许多研究人员交流思想，沟通情况，这相当于开一个小型电子会议。超文本信息资源是按知识单元及其关系建立的知识结构网络。其数据库由节点和链路组成，查阅超文本信息资源时，以知识片段及其关系作为追踪、检索的依据。除了处理一般的文字信息外，还包括图片、地图和其他直观信息。超文本信息资源能够把文字信息和图像信息有机地结合在一起。

③ 多媒体信息资源：多媒体是指包括文本、图像和声音在内的各种信息或传播形式的总称。多媒体信息能针对用户的需求提供各种形式的信息。它们可以是文本、图像（图表、图画、照片、动画或活动影视）和声音（语言、音乐或其他音响）以及它们的结合。由于计算机软、硬件技术的限制，相当长时间以来计算机信息检索系统只限于存储和检索书目、文摘等线索型文献，多媒体的出现使得人们接受的信息资源不但图、文、声并茂，而且丰富多彩。

④ 超媒体信息资源：超媒体则是超文本与多媒体两种技术的结合。一般说来，当超文本节点中的信息是多媒体信息时，即在信息浏览环境下超文本的信息管理方式与多媒体的信息表现方法结合在一起时，就称为超媒体，它是超级媒体的简称。近几年来，超媒体技术发展迅速，在 Internet 上超媒体应用系统不断涌现。在超媒体信息系统中，不同类型的媒体信息能高度综合和集成。空间上，图、文、声并茂；时间上，媒体信息同步实现，有超文本和多媒体两种信息资源的特点，具有高度的交互性。

（3）按照信息的交流方式划分。按人类信息交流的方式，数字信息资源分为非正式出版信息、半正式出版信息和正式出版信息 3 种。

① 非正式出版信息：非正式出版信息包括电子邮件、网络论坛、电子会议和电子布告版新闻等。这类信息流动性、随意性强，信息质量难以控制和保证。

② 半正式出版信息：半正式出版信息包括内部电子期刊、会议文集、各类报告、机构情况和产品简介等。这类信息内有的纳入正式出版信息系统，受到一定的知识产权保护。

③ 正式出版信息：正式出版信息按加工层次可划分为一次出版信息（如电子图书、电子期刊、电子报纸等）、二次出版信息（如检索数据库、搜索引擎、网络导航等）和三次出版信息（如参考数据库、网络书评等）。

2. 数字信息资源的特点

（1）信息存储形式分为文本→超文本→多媒体→超媒体。这使得信息的组织方式发生了巨大的变化，不仅以知识的信息为基本单元，而且充分展示了这些单元之间的逻辑关系，为网络环境下不同形式的信息资源的管理和开发利用提供了支持。由传统的顺序、线性排列，通过利用数字化存储技术，发展到超文本、超媒体技术，使得信息可按照自身的逻辑关系组成相互联系的、直接的和非线性的网状结构。

（2）存储介质发生转换。信息资源由纸张上的文字变成磁性介质上的电磁信号或光介质上的光信号，从模拟信息转变为数字信号，使信息的存储传递和查询更加方便，且存储信息密度高、容量大，可以无损耗地被重复利用。

（3）以现代信息技术为记录手段，是一种数字化的信息资源。信息以数字化的形式存在，

既可在计算机内高速处理，又可借助通信网络进行远距离传播，这使得共享全球信息资源成为可能。而且随着网络的进一步扩大与应用范围的拓宽，数字化的信息资源将成为信息资源的最终转化方式。

(4)内容丰富。它既可以是文字、图表等静态信息，也可以是集图、文、声、像于一体的动态多媒体信息，并且各种类型的数据又可借助计算机实现任意的组合编辑；把枯燥的文字信息转化为形式多样、活泼的数字信息，界面友好，易于人机沟通。

(5)数据结构具有通用性、开放性和标准化的特点。在网络环境下，可被数人同时访问，是一种共享性的信息资源。在讲求兼容性与标准化的信息社会中，数字信息资源易于实现信息资源的扩充，以及各信息之间的互联与互操作。

(6)具有高度的整合性。数字信息资源不受时间、空间的限制，可以实现跨时空、跨行业的传播。

(7)便于各种媒介信息的一体化。数字信息资源的相互转换和二次开发易于形成各种数据库，便于检索与使用，提高了信息资源的利用价值。

(8)交互式性能增强。由于数字信息资源存储在计算机能够识别的介质上，因此随着计算机软件的更新与性能的日益提高，用户逐渐具有更多的主动性。他们不仅是数字信息资源的利用者，而且将成为数字信息资源的开发主体。

第二节　信息检索

一、信息检索概念

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信息的过程。

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、准确、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。

二、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，即对信息集合与需求集合的匹配与选择。

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

传统信息检索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)对于涉及几个概念组合的多主题的文献难于找到。

2. 现代信息检索

现代信息检索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、计算机技术和网络通信技术等。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,并向着智能化的方向发展。

三、信息检索研究对象

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储与检索有关的系统、过程、理论和方法。一切可供存储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、存储、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

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、演变以及设计开发技术、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、存储技术革新、检索技术、系统管理和经营知识及市场营销技术等,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。

第三节　检索语言

一、检索语言的概念

检索语言是根据信息存储与检索的需要而创造的一种人工语言。检索语言是信息检索与信息存储的一种约定语言。

检索语言与检索效果之间有着密切的关系，它在检索过程中起着极其重要的作用。信息检索的全过程包括信息的存储过程和检索过程。当存储信息时，文献标引人员首先要对各种文献进行主题分析，通过分析选出若干个能代表文献主题的概念，并用检索语言把这些概念标引出来，然后纳入信息检索系统中。当检索信息时，信息检索人员首先也要对检索课题进行分析，并且通过分析明确其检索范围，选出若干个能代表信息需要的概念，并把这些概念转换成检索语言，然后从信息检索系统中查找用该检索语言标引的文献，从而找到用户所需的信息。

由此可见，检索语言是信息检索系统的主要组成部分，是标引人员与检索人员之间沟通思想、取得一致意见的桥梁。

二、检索语言的组成

实质上，信息检索语言是表达、概括文献信息内容的概念及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分类号码，又可以是代表某类事物的某一方面特征的一套代码（如化合物的各种代码），用以对文献内容和信息需要进行主题标引、逻辑分类或特征描述。

检索语言是由词汇和语法组成的。在这里，词汇指的是登录在分类表、词表中的全部标识，一个标识（分类号、检索词、代码）就是它的一个语词，而分类表及词表则是它的词典。语法指的是如何创造和运用标识（单个标识或几个标识的组合）来正确表达文献内容和信息需要，以有效地实现信息检索的一整套规则。

三、检索语言的种类

检索语言按其反映信息内外部特征的不同可分为：分类语言、主题语言、名称语言和代码语言4大类。

1. 分类语言

分类语言是一种直接体现知识分类的等级结构的标识系统。它根据一定的观点，以科学分类为基础，以文献内容的科学性质为对象，运用概念划分与概括的方法，按照知识门类的逻辑次序，从一般到具体，从简单到复杂，进行层层划分；每划分一次，就产生许多类目；逐级划分，就产生许多不同级别的类目。所有不同级别的类目，层层隶属，形成一个严格有序直线性的知识门类的等级制体系。每个类目都用分类符号作为标记，每个分类号都是表达特定知识概念的词汇，这些词汇即是分类语言。

分类语言是用分类号码来表达各种概念，将各种概念按学科性质进行分类和系统排列，便于信息存储与信息检索双方进行交流的一种检索语言。著名的《国际十进分类法》、《美国国会图书馆图书分类法》、《中国图书馆图书分类法》等，即是以分类语言为依据广泛应用于信息存储与信息检索的规范，是对信息按学科属性及技术特点进行有序化和检索利用的重要工具。

分类语言具有按学科或专业集中、系统地揭示信息内容的功能，有利于从学科或专业角度进行全面地检索；按着结构逐级划分，具有等级结构，便于扩大和缩小检索范围。

2. 主题语言

主题语言是用语词来表达各种概念，将各种概念按字顺排列。主题语言包括标题词语言、关键词语言、单元词语言和叙词语言等。它们统称为主题法系统。

(1)标题词语言是规范化了的自然语言。它以经过标准化处理的名词术语作为标识，来直接表达文献所论及或涉及的事物——主题之间的相互关系(这种关系是借助于参照系统来间接显示的)。

(2)关键词语言是为适应主题目录及主题索引编制自动化的需要而产生的一种主题语言型检索语言。关键词是指在文献的标题、摘要或正文中出现的，对表达文献内容具有实质意义，能作为检索入口的，起关键性描述作用的词汇；是针对文献中的关键词选定或抽出，极少用做词汇控制，按字顺排列，从而提供检索途径的一种检索语言。

(3)单元词语言又称为元词语言，它是从文献中抽取出来并经过控制处理的，能表达文献主题的最小、最基本的词汇单位。它可以是一个单纯词，也可以是一个合成词。这些词具有一个共同的特点：它们在概念上不能再进一步分解，如再分解，就再也不能表达原来所代表的特定概念，从而失去检索本意。

(4)叙词语言是经过词汇控制的，在标引中用来显示文献主题，在检索中用来构成表达式的一种检索语言。叙词语言可谓是博采各种信息检索语言之长，吸取了多种信息检索语言的原理和方法。叙词语言是一种采用规范化的单词或词组由标引人员或读者自行组配，以表达文献(或课题)主题概念的一种后组式索引语言，或者叫做后组式的检索语言。

3. 名称语言

名称语言是以人名、机构名、地名、书名、刊名、篇名等代表信息特征的名称为检索标识，作为标引文献和检索文献双方共同采用的交流语言。各种数据库中所设置的作者检索途径、机构检索途径、出版物检索途径等都是运用名称语言对信息的特征予以描述和展示的结果。

4. 代码语言

代码语言一般只就事物的某一方面特征，用某种代码系统来加以标引和排列，如专利号、标准号、化学物质登记号等。

四、检索语言的功能

检索语言不同于自然语言，它所表述的概念只有一种解释，不允许一词多义，多词一义而使概念的表述模糊不清。检索语言的这种单义性保证了表述概念的惟一性及标引与检索的一致性，从而使信息检索人员又全、又准、又快地检索到含有所需信息的文献。

检索语言的功能是通过检索语言标引文献的主题概念，不仅能简明地提示文献所包含的信息内容及其外表特征，而且还能将同一主题概念的文献集中在一起，使文献的存储集中化、系统化、组织化，以便于进行有规律的检索。

第四节　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、存储、选择和传播等功能。

一、信息检索系统的类型

1. 按照信息存储与检索使用的设备划分

其可分为手工信息检索系统和计算机信息检索系统两种类型。

(1)手工信息检索系统。手工信息检索系统是在计算机信息检索系统出现以前进行信息检索的主要检索工具，主要包括书本式信息检索系统和卡片式信息检索系统。

① 书本式信息检索系统：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。书本式信息检索系统编制原理是计算机信息检索系统产生的基础。

② 卡片式信息检索系统：它包括一切以普通卡片存储和查找信息的工具，如图书馆内部的各种卡片目录，管理部门的各种卡片档案。与书本式检索系统相比，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，其正在逐渐被计算机目录系统所取代。

(2)计算机信息检索系统。计算机信息检索系统主要由计算机硬件和软件系统、数据库、数据通信等设备组成。根据内容不同，计算机信息检索系统分为：联机检索系统、光盘检索系统、网络检索系统。

① 联机检索系统：国际联机检索就是用户使用终端设备，远距离地从国际联机检索中心迅速而准确地获取数字文献信息资源，使知识信息得到广泛而有效地传播和利用；其实质是数据库和通信的结合。从20世纪六七十年代起，许多国家还先后建立了专门从事计算机信息检索的机构，如美国的洛克希德公司和系统发展公司，英国的目录检索服务处和图书馆自动化情报服务处与Infoline公司以及欧洲的ESA等，这些机构都建有大量的数据库联机检索系统，都向全世界联机用户提供电子信息服务。其中，著名的系统有Dialog，ORBIT，ESA－IRS，OCLC，STN等。

② 光盘检索系统：由于光盘在存储电子信息资源方面具有记录密度高、容量大、成本低、体积小、寿命长、可实现随机存取和检索费用低廉等优点，因此光盘被广泛用于存储、检索数字信息资源，并产生了一批生产系列光盘的公司，如美国UMI公司和银盘公司等。光盘记载的数字信息资源并不局限于文献信息，还包括各种软件，但可用于检索，仍以文献信息为主。

③ 网络检索系统：Internet是网络通过互联而形成的全球网。它已延伸到地球上几乎每个国家。在Internet网上的所有主机都采用TCP/IP协议连接和通信，使网上各种计算机都遵循该协议所规定的方式进行数据交换，其结果是使得Internet信息资源囊括了电子报刊、电子新闻、电子报告、电子论坛、会议资料、各种软件资料、图像文件、声音文件和电子游戏等。Internet是目前世界上资料最多、门类最全、规模最大的信息库，是人们获取信息的重要来源。

2. 按照揭示信息的内容程度划分

其可分为目录、题录、文摘和全文数据库。