

信息检索

XINXI JIANSUO
YU LIYONG

与利用



主 编 曾英姿

副主编

罗 宏 李红霞 何晓庆
杨 莹 刘裴裴 张 宸



四川大学出版社
SICHUAN UNIVERSITY PRESS

信息检索

INXI JIANSUO
YU LIYONG

与利用



主 编 曾英姿
副主编 罗 宏 李红霞 何晓庆
杨 莹 刘裴裴 张 宸



四川大学出版社
SICHUAN UNIVERSITY PRESS

特约编辑:梁 平
责任编辑:楼 晓
责任校对:武慧智
封面设计:李金兰
责任印制:王 炜

图书在版编目(CIP)数据

信息检索与利用 / 曾英姿主编. —成都: 四川大学出版社, 2012. 11
ISBN 978-7-5614-6265-2

I. ①信… II. ①曾… III. ①情报检索 IV.
①G252.7

中国版本图书馆 CIP 数据核字 (2012) 第 264943 号

书名 信息检索与利用

主 编	曾英姿
出 版	四川大学出版社
地 址	成都市一环路南一段 24 号 (610065)
发 行	四川大学出版社
书 号	ISBN 978-7-5614-6265-2
印 刷	郫县犀浦印刷厂
成品尺寸	185 mm×260 mm
印 张	14.25
字 数	361 千字
版 次	2013 年 1 月第 1 版
印 次	2013 年 1 月第 1 次印刷
定 价	29.00 元

版权所有◆侵权必究

◆读者邮购本书,请与本社发行科联系。电话:85408408/85401670/85408023 邮政编码:610065

◆本社图书如有印装质量问题,请寄回出版社调换。

◆网址: <http://www.scup.cn>

前 言

当今世界一个非常显著的时代特征：全球性信息化程度迅猛提升、多元信息资源与噪音无限膨胀、信息掌控与利用需求日益彰显，人类社会正在加速迈入以社会知识化、信息网络化、经济全球化、教育终身化、学习社会化等为特征的知识经济时代。信息业已凌驾于资源、能源等基本社会要素之上，成为掌控人类工作、学习、生活、娱乐的核心要件。对信息的采集、加工和利用能力，将成为个人、团体、地区乃至国家的核心竞争能力。掌握信息获取与利用方法，已成为社会所有成员必须具备的基本技能。

信息检索 (information retrieval, IR) 作为一门专门研究信息存储与信息获取的学科，乃顺应信息化浪潮的新兴综合性研究领域。20 年前，人们还主要通过他人而非亲手检索来获取信息。然而，近年随着计算机、网络、通讯等技术的迅猛发展，信息载体和传播途径的巨大变化，除传统纸质信息外，大量磁载体信息、电子版信息及各类网上信息呈海量涌现。这些浩如烟海的信息具有多样性、离散性、无序性、重复性等特征，极大地影响了人们获取信息的质量与效率。基于前述时代背景，信息检索对于培养和提高人们的信息获取能力、信息组织能力、信息综合能力和信息评价能力，促进个体与社会团体核心竞争能力的不断提升，具有十分深远的意义。相关研究发现，美国等信息化程度较高的国家，信息检索已经从一个以学术研究为主的领域，变成了人们赖以获取日常信息的工具性学科。《礼记·学记》云：“善学者，师逸而功倍，又从而庸之；不善学者，师勤而功半，又从而怨之。”古人治学讲究善学善问，方能事半功倍；现今信息时代，善于发现、捕捉与利用信息，即信息素养应当成为当代大学生必备的基本素质。信息素质教育必将成为高等教育的重要使命之一。

信息检索是信息资源管理学科的热点领域之一。其发展主要包括如下几个方面：首先，检索手段与平台的快速更新换代。当今手工检索方式基本淘汰，取而代之的是基于因特网环境的计算机信息检索。其次，检索技术也得到飞速发展，不再是布尔检索一统天下，而是一个集布尔检索、全文检索、超文本检索、智能检索等多种检索技术于一体的综合检索。伴随着信息检索原理与技术研究的不断深入，国内外学术界涌现了不少优秀的学术成果。再次，检索对象得以快速延伸。信息检索的对象已由科学文献和馆藏记录扩展到其他类型的内容，特别是记者、律师、医生等特定领域专业人士所需的信息内容。此外，随着因特网的发展及其应用的普及，信息检索理论不断拓展、延伸和丰富。

本书是编者多年来追踪信息检索领域发展动向、开展信息检索相关问题的研究与教学的经验总结，对信息检索领域的相关问题进行了简明扼要的梳理，既有对其基本理论、方法、技术的论述，也有对其最新发展的系统阐述。本书以信息检索的基本原理和科学方法为主线，精选目前最新的具有代表性和权威性的文科和理科的主要数据库为对象，以清晰

的条理和大量的实例,全面、系统地介绍了各类网络数据库的基本特征、信息价值、编制体例、检索方法和检索技巧等方面的知识。本书还结合我国高等教育信息保障体系的特点,介绍了信息保障体系的相关知识(CALIS、CASHL、NSTL、国家图书馆等),使读者能够较为全面地了解并掌握现代网络条件下的信息资源交流渠道和利用方法,全方位地提高和扩展信息检索和利用的能力。另外,还引入了个人文献管理软件(NoteExpress、EndNote等)、科研评价的引文数据库(CSSCI、CSCD、CCD、DISC、JSR等)以及学科、学校科研的评价系统(ESI)等时下信息应用领域的热门内容。最后还介绍了论文的写作格式。全书组织结构较为简明,既可用来系统学习,也能够作为独立篇章进行查询。

本书共10章,全书的体系结构由曾英姿构思,并经全体编者共同讨论确定,曾英姿负责全书的修改、统稿。各章节的执笔人分别是:曾英姿(前言、体系大纲、体例制订)、罗宏(第4章4.1、4.4、4.5节,第5章5.3节,第6章)、李红霞(第5章5.2节、第8章)、何晓庆(第7章、第9章)、杨莹(第1章、第2章、第3章3.1~3.7节)、刘裴裴(第4章4.3节、4.6~4.9节,第5章5.5、5.6节,第10章)、张宸(第3章3.8、3.9节,第4章4.2、4.10~4.13节,第5章5.1、5.4节)。

本书内容新颖、材料丰富、深入浅出,可以作为高年级本科生和研究生的课程教材。本书注重理论与实践的有机结合,理论方面侧重于论述信息检索的原理、方法及其支撑技术,实践部分侧重于阐述信息资源及其检索工具和方法,并有大量实例供参考。

本书编写期间得到了参与单位的领导和教师们的大力支持,谨致衷心感谢!最后,我们还要感谢书稿的审阅人员,他们对本书的内容和结构提出了宝贵建议。我们对他们的指导表示深深的谢意!

由于编者学识与水平有限,书中不足之处在所难免,欢迎广大读者和同行、专家们批评指正。

编 者

目 录

第 1 章 信息检索基础知识和理论	(1)
1.1 信息检索基本概念	(1)
1.2 文献信息资源的类型与识别	(3)
1.3 信息检索的语言、方法和途径	(6)
1.4 计算机检索基础知识及检索技术	(11)
第 2 章 常用中文数据库检索	(16)
2.1 中国学术文献网络出版总库 (CNKI)	(16)
2.2 中文科技期刊数据库 (维普)	(19)
2.3 万方数据知识服务平台	(22)
2.4 超星数字图书馆	(25)
2.5 读秀图书搜索	(25)
2.6 方正 Apabi 数字图书馆	(26)
2.7 书生数字图书馆	(27)
第 3 章 文科类中文数据库检索	(29)
3.1 全国报刊索引数据库	(29)
3.2 人大复印报刊资料数据库	(30)
3.3 龙源电子期刊	(32)
3.4 北大法宝	(32)
3.5 中国基本古籍数据库	(36)
3.6 国务院发展研究中心信息网 (国研网)	(37)
3.7 中经专网	(38)
3.8 EPS 全球统计数据/分析平台	(40)
3.9 国泰安数据库	(44)
第 4 章 国外综合性检索工具	(47)
4.1 Encyclopedia Britannica Online (不列颠百科全书)	(47)
4.2 MyiLibrary 电子书	(50)

4.3	KingBook (金图国际外文数字图书馆)	(52)
4.4	EBSCO	(54)
4.5	OCLC FirstSearch	(59)
4.6	Oxford Journals (OUP)	(62)
4.7	ProQuest Research Library (PRL)	(65)
4.8	SAGE Journals Online (SJO)	(68)
4.9	ProQuest (PQDT) 学位论文全文数据库	(71)
4.10	SpringerLink	(73)
4.11	Science Direct (Elsevier SDOL)	(75)
4.12	Wiley Online Library	(77)
第 5 章	国外理科类专业性检索工具	(81)
5.1	BIOSIS Previews (OvidSP) (生物学文摘)	(81)
5.2	SciFinder Scholar (CA 网络版) (化学文摘)	(86)
5.3	Ei Compendex (工程索引)	(92)
5.4	Inspec (ISI) (科学文摘)	(96)
5.5	MathSciNet (数学评论)	(102)
5.6	全文电子期刊	(105)
第 6 章	国外文科类专业性检索工具	(111)
6.1	Gale 参考数据库	(111)
6.2	EEBO (早期英文图书在线)	(116)
6.3	JSTOR (Journal Storage)	(118)
6.4	PAO (典藏学期刊全文数据库)	(120)
6.5	Westlaw	(122)
6.6	Lexis.com	(126)
6.7	Emerald	(130)
6.8	East View Universal Database (俄罗斯大全数据库)	(132)
第 7 章	专利及专利信息的检索	(134)
7.1	专利概述	(134)
7.2	国家知识产权局专利数据库	(139)
7.3	Derwent Innovation Index	(142)
7.4	美国专利	(145)
7.5	欧洲专利	(148)
7.6	加拿大专利	(150)

第 8 章 引文数据库	(153)
8.1 概况	(153)
8.2 中文社会科学引文索引 (CSSCI)	(155)
8.3 中国科学引文数据库 (CSCD)	(157)
8.4 中国引文数据库 (CCD)	(160)
8.5 国际科学引文数据库 (DISC)	(162)
8.6 Web of Science	(164)
8.7 Journal Citation Reports (JCR)	(174)
8.8 Essential Science Indicators (ESI)	(176)
第 9 章 国内外信息资源服务保障体系	(178)
9.1 CALIS (中国高等教育文献保障系统)	(178)
9.2 CASHL (中国高校人文社会科学文献中心)	(181)
9.3 国家图书馆	(184)
9.4 上海图书馆上海科技情报所	(188)
9.5 NSTL (国家科技图书文献中心)	(189)
9.6 CADAL (大学数字图书馆国际合作计划)	(192)
9.7 OCLC	(194)
9.8 OhioLink	(195)
9.9 ELIB	(196)
9.10 BLDSC	(197)
第 10 章 NoteExpress、EndNote 文献管理与论文写作格式	(199)
10.1 NoteExpress	(199)
10.2 EndNote	(204)
10.3 论文写作格式	(211)
参考文献	(214)

第1章 信息检索基础知识和理论

1.1 信息检索基本概念

1.1.1 信息、知识、文献的概念

1. 信息

信息又称讯息、资讯，是信息论中的一个术语。信息论的创始人香农认为：信息是能够用来消除不确定性的东西。至今信息还没有一个公认的定义，不同研究领域的专家学者从各自的学科角度出发，对于信息定义有不同的描述。但是深刻理解信息的本质特征后，可以用“记录”来说明信息的本质，信息是事物在相互作用中所刻画出的记录。

在最一般的意义上，可以将信息定义为：信息是世界上一切事物的状态和特征的反映，是用文字、数据或信号等形式，通过一定的传递和处理，来表现各种相互联系的客观事物在运动变化中所具有特征内容的总称。这里的“事物”泛指存在于人类社会、思维活动和自然界中一切可能的对象。“运动变化”则是指事物在时间和空间上变化信息的载体所展示的特征、态势和规律。

2. 知识

人类的认识成果来自社会实践。其初级形态是经验知识，高级形态是系统科学理论。按其获得方式可分为直接知识和间接知识。按其内容可分为自然科学知识、社会科学知识、思维科学知识和哲学知识。其中哲学知识是关于自然、社会和思维知识的概括和总结。知识的总体在社会实践的世代延续中不断积累和发展。

知识是抽象的，是传达概念的一种形式。例如：古代结绳记事、白旗示降等，都是借由结绳、白旗等符号来传达概念。但是当人类活动越来越频繁后，简单的符号就不足传达双方的意念了。因此，有了文字、语言等符号兴起以作为沟通的工具。在沟通过程中还需要经过思考，最后才能产生出知识。概括来说，知识是人类社会实践的总结，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和掌握，是人脑通过思维重新组合的系统化的信息集合。

3. 文献

文献是用文字、图形、符号、声频、视频等技术手段记录人类知识的一种载体，可理解为固化在一定物质载体上的知识，也可以理解为古今一切社会史料的总称。现在通常认为文献为图书、期刊等各种出版物的总和。文献是记录、积累、传播和继承知识的最有效手段，是人类社会活动中获取情报的最基本、最主要的来源，也是交流传播情报的最基本

手段。

1984 年的中华人民共和国国家标准《文献著录总则》关于“文献”的定义是：“文献：记录有知识的一切载体。”在这一定义中，有两个关键词：“知识”，是文献的核心内容；“载体”，是知识赖以保存的物质外壳，即可供记录知识的某些人工固态附着物。也就是说，除书籍、期刊等出版物外，凡载有文字的甲骨、金石、简帛、拓本、图谱乃至缩微胶片、视盘、声像资料等等，皆属文献的范畴。

文献具有三个基本属性：知识性、记录性和物质性。文献具有存贮知识、传递和交流信息的功能。

文献构成的基本要素包括：知识内容、记录符号、载体材料、制作方式和载体形态。例如：易中天所著《品三国》这本专著文献，书中所论述的内容就是知识内容，著书所用的汉字就是记录符号，纸张是载体材料，印刷是制作方式，纸质书就是载体形态。从以上分析可以看出文献既具有精神属性，也具有物质属性，它是一定知识内容和一定物质形式的统一体。

4. 信息、知识、文献之间的关系

信息是事物的本来面貌，信息经过人脑加工形成知识。信息与知识密不可分，信息是构成知识的原料，知识的产生离不开信息。信息是广泛存在于自然界和人类社会中的一切事物的存在方式和运动状态的客观反映，是人类认识世界的依据；知识是人类创造的精神财富，是人类接受了来自自然界和社会中的大量信息后，将反映自然现象和社会现象的信息经过加工，上升为对自然界和人类社会客观规律的认识；信息、知识被记载在载体上就形成文献。文献是记录有知识和信息的一切载体，是知识或信息的重要存储和传播工具。文献经过传递、传播、应用于理论与实际而产生新的信息。总之，信息、知识与文献之间是相互包含、相互转换与相互再创造的关系。

1.1.2 信息检索的定义

信息检索又叫文献信息检索或情报检索，是指将信息按一定的方式组织和存贮起来，形成各种信息集合（通常称为检索工具或检索系统），然后根据用户的需要，按照一定的检索途径和方法，从信息库中找出符合用户需要的信息的过程。因此，信息检索的全过程包括信息标引和存储过程以及信息的检索过程。而狭义的信息检索就是信息的检索过程，即从信息集合中找出所需要的信息的过程。

在信息的存贮过程中，信息标引人员或自动标引程序对各种信息进行主题分析，将信息所包含的内容分析出来，形成多个能代表信息主题的概念，并用信息检索语言的词语把这些概念表达出来，归入检索工具；而在信息检索过程中，检索者首先对信息需求进行主题分析，形成能代表检索需求的概念，并将这些概念转换成信息检索语言的词语，然后在检索工具中进行匹配运算，从而找到所需的信息。因此，信息检索的实质是一个匹配过程，是将描述特定用户所需信息的提问特征，与信息存储的检索标识进行异同的比较，从中找出与提问特征一致或基本一致的信息。这里的提问特征是指欲检索课题中能代表信息需求的主题词、关键词或分类号等。

1.1.3 信息检索方式

根据信息检索手段的不同,信息检索方式分为手工检索和计算机检索。

1. 手工检索 (简称手检)

手工检索是指用人工方式查找所需信息的检索方式。检索的对象是书本型的检索工具,检索过程由人脑和手工操作相配合完成,匹配的过程是人脑的思考、比较和选择的过程。手工检索便于控制检索的准确性,查准率较高,但是检索过程长,速度慢,查全率较低。

2. 计算机检索 (简称机检)

计算机检索是指把信息及其检索标识转换成电子计算机可以阅读的二进制编码,存储在磁性载体上,由计算机根据程序进行查找和输出。

计算机检索的对象是计算机检索系统,针对数据库进行,检索过程在人与计算机的协同作用下完成,匹配是由机器完成的。与手检相比,机检检索的本质没有改变,变化了的只是信息的媒体形式、存储方式和匹配方法。计算机检索查全率较高,查准率较低。

随着检索技术的发展,计算机检索又分为脱机检索、联机检索、光盘检索和网络检索。

(1) 脱机检索:成批处理检索提问的计算机检索方式。

(2) 联机检索:检索者通过检索终端和通信线路,直接查询检索系统数据库的机检方式。Dialog 国际联机情报检索系统是世界著名的联机检索系统之一。

(3) 光盘检索:以光盘数据库为基础的一种独立的计算机检索,包括单机光盘检索和光盘网络检索两种类型。

(4) 网络检索:利用 Email、FTP、Telnet、Archie、WAIS、Gopher、Veronica、WWW 等检索工具,在 Internet 等网络上进行信息存取的行为。

随着社会信息化、网络化进程的发展,信息检索已全面进入网络检索时代:一方面,基于 WWW 环境的各种搜索引擎系统发展迅速;另一方面,传统的联机检索服务系统和各种数据库检索系统也纷纷将服务平台转移到了网络中来。

1.2 文献信息资源的类型与识别

以文字、图形、符号、声频、视频等方式记录在各种载体上的知识和信息资源,统称为文献信息资源。它记录着无数有用的事实、数据、理论、方法、假说、经验和教训,是人类进行交流、认识和改造世界的基本工具。这类信息由于经过加工、整理,故较为系统、准确、可靠,便于保存与利用。

文献信息资源的划分有三种方式:根据文献信息的载体形式划分、根据文献信息的编辑出版形式划分和按文献信息的结构等级(即文献信息的加工程度)划分。

1.2.1 根据文献信息的载体形式划分

根据载体形式的不同把文献信息分为印刷型文献、缩微型文献、机读型文献和声像型

文献。

印刷型文献是最基本形式的文献,包括铅印、油印、胶印、石印等各种资料,其优点是可直接、方便地阅读。

缩微型文献是以感光材料为载体的文献,又可分为缩微胶卷和缩微平片,优点是体积小,便于保存、转移和传递。但阅读时须用阅读器,阅读的范围受到限制。

机读型文献(计算机阅读型)是一种最新形式载体的文献。它主要通过编码和程序设计,把文献变成符号和机器语言,输入计算机,存储在磁带或磁盘上,阅读时,再由计算机输出,它能存储大量情报,可按任何形式组织这些情报,并能以极快的速度从中取出所需的情报。电子图书即属于这种类型。

声像型文献又称直感型或视听型文献,是以声音和图像形式记录在载体上的文献,如唱片、录音带、录像带、科技电影、幻灯片等。

1.2.2 根据文献信息的编辑出版形式划分

根据文献信息的编辑出版形式,信息资源可以划分为图书、期刊(连续出版物)、学位论文、会议文献、专利文献、科技报告、标准文献、政府出版物、档案资料、产品资料等。在学术研究中,学术价值和使用频率较高的是前几种文献类型。

1. 图书

图书是人类社会实践的产物,是一种特定的不断发展着的知识传播工具。联合国教科文组织对图书的定义是:凡由出版社(商)出版的不包括封面和封底在内 49 页以上的印刷品,具有特定的书名和著者名,编有国际标准书号,有定价并取得版权保护的出版物称为图书。

联合国教科文组织在对图书的定义中指出图书编有国际标准书号。国际标准书号(International Standard Book Number)简称 ISBN,是国际通用的图书或独立的出版物(除定期出版的期刊)代码。出版社可以通过国际标准书号清晰地辨认所有非期刊书籍。一个国际标准书号只有一个或一份相应的出版物与之对应。

2. 期刊

期刊也称杂志,亦称连续出版物,是指有固定名称,用卷、期或者年、季、月顺序编号,按照一定周期出版的成册连续出版物。

期刊按用途不同可以分为科普类期刊和学术类期刊两大类。学术期刊按主管单位的不同,又可以分为国家级、省部级和地市级;按所发表文章学术水平高低分为核心期刊和非核心期刊,其中中文核心期刊包括中文核心期刊要目总览(北大版)、中文社会科学引文索引数据库(CSSCI)和中国科学引文数据库(CSCD)。

期刊还可以按照科目分类,如社会科学、哲学、经济学、法学、教育学、文学、历史学、自然科学、理学、工学、农学、医学等。

根据期刊的出版周期可将期刊分为周刊、半月刊、月刊、双月刊、季刊、半年刊和年刊。

每种期刊有固定的 ISSN 号(国际标准连续出版物编号,International Standard Serial Number),是根据国际标准制定的连续出版物国际标准编码,其目的是使世界上每一种不

同题名、不同版本的连续出版物都有一个国际性的唯一代码标识。

3. 学位论文

学位论文是指高等学校或研究院所的学生在导师指导下从事某一学术课题的研究,为获得某种学位而撰写的学术性较强的研究论文。这类论文在格式和内容等方面有严格要求,学位论文是学术论文的一种形式。

根据《中华人民共和国学位条例》的规定,按照所获学位级别的不同,学位论文分为学士论文、硕士论文、博士论文三种。

按照研究方法不同,学位论文可分理论型、实验型、描述型三类。理论型论文运用的研究方法是理论证明、理论分析、数学推理,并用这些研究方法获得科研成果;实验型论文运用实验方法,进行实验研究获得科研成果;描述型论文运用描述、比较、说明方法,对新发现的事物或现象进行研究而获得科研成果。

4. 会议文献

会议文献是在各种学术、专题会议上发表的论文和报告。会议文献多数以会议录的形式出现,是报道专业领域最新研究成果的一种主要方式。随着科学技术的迅速发展,世界各国的学会、协会、研究机构及国际性学术组织举办的各种学术会议日益增多。

会议文献没有固定的出版形式,有的刊载在学会协会的期刊上,作为专号、特辑或增刊,有些则发表在专门刊载会议录或会议论文摘要的期刊上。一些会议文献还常常汇编成专题论文集或出版会议丛刊、丛书。还有些会议文献以科技报告的形式出版。

5. 专利文献

专利文献是申请或批准专利的发明创造,即包含已经申请或被确认为发现、发明、实用新型和工业品外观设计的研究、设计、开发和试验成果的有关资料,以及保护发明人、专利所有人及工业品外观设计和实用新型注册证书持有人权利的有关资料的已出版或未出版的文件(或其摘要)的总称。这些资料是在专利审批过程中产生的文件,有些是公开出版的,有些则仅为存档或仅供复制使用。作为公开出版物的专利文献主要包括:专利申请说明书、专利说明书、实用新型说明书、工业品外观设计说明书、专利公报、专利索引等。

6. 科技报告

科技报告是在科研活动的各个阶段,由科技人员按照有关规定和格式撰写的,以积累、传播和交流为目的,能完整而真实地反映其所从事科研活动的技术内容和经验的特种文献。

7. 标准文献

标准文献是指标准化工作的文件,是经过公认的权威当局批准的标准化工作成果,主要为有关工业产品和工程建设的质量、规格和检验方法的技术规定文件。作为一种规章性文献,它具有一定的法律约束力。一个国家的标准文献反映该国的生产工艺水平和技术经济政策,而国际现行标准则代表了当前世界水平。国际标准和工业先进国家的标准常是科研生产活动的重要依据和信息来源。

1.2.3 根据文献信息的结构等级(文献信息的加工程度)划分

按文献信息的结构等级,即对文献信息的加工程度可以将文献划分为零次文献、一次

文献、二次文献和三次文献。

零次文献：指未经过任何加工的原始文献，如实验记录、手稿、原始录音、原始录像、谈话记录等。零次文献在原始文献的保存、原始数据的核对、原始构思的核定（权利人）等方面有着重要的作用。零次文献具有一定的学术、史料价值。非常重要的零次信息资源会作档案保存或作数据备份保存。

一次文献：指作者以本人的研究成果为基本素材而创作或撰写的文献，不管创作时是否参考或引用了他人的著作，也不管该文献以何种物质形式出现。专著、期刊论文、学位论文、会议文献、专利文献以及科技报告等文献均属于一次文献。

二次文献：指文献工作者对一次文献进行加工、提炼和压缩之后所得到的产物，是为了便于管理和利用一次文献而编辑、出版和累积起来的工具性文献。检索工具（包括目录、题录、索引、文摘、数据库等）和网上检索引擎是典型的二次文献。

三次文献：指对有关的一次文献和二次文献进行广泛深入的分析研究、综合概括而成的产物，如教科书、大百科全书、辞典、手册、年鉴、综述等。

一般来说，一次文献是基础，是最基本的信息源，是检索利用的对象；二次文献是一次文献的集中提炼和有序化，是文献信息检索和利用的主要对象，也被称之为检索工具；三次文献是把分散的一、二次文献按专题或知识的门类进行综合分析研究加工而成的成果，是高度浓缩的文献信息，它既是文献信息检索利用的对象，又可作为检索文献信息的工具。

1.2.4 文献信息出版类型识别

各种出版类型的文献有其固定的书写格式。有正确的描述即著录，才能有正确的识别。2005年3月，国家质量监督检验检疫总局和中国国家标准化委员会发布了GB/T 7714—2005《文后参考文献著录规则》。该标准规定了各个学科、各种类型出版物的文后参考文献的著录项目、著录顺序、著录用的符号、各个著录项目的著录方法以及参考文献在正文中的标注法。它是专供著者和编辑使用的文献著录规则。各种出版类型的文献的著录格式均应遵循《文后参考文献著录规则》的规定，格式如下：

- (1) 图书：作者. 书名 [M]. 版次（第1版不写）. 出版地：出版者，出版年：起止页.
- (2) 期刊：作者. 题名 [J]. 刊名，年，卷（期）：起止页.
- (3) 学位论文：作者. 论文题名 [××学位论文]. 授予单位所在地：授予单位，授予年.
- (4) 会议论文：姓名. 题名 [C]. 会议录名：起止页码，年.
- (5) 专利文献：专利申请者. 专利题名：专利国别，专利号 [P]. 发布日期.
- (6) 科技报告：著者. 题名 [R]. 报告题名，编号，出版地：出版者，出版年：起止页码.
- (7) 标准文献：标准代号. 标准名称 [S]. 出版地：出版者，出版年.

1.3 信息检索的语言、方法和途径

1.3.1 信息检索的语言

计算机检索基本原理是将用户的检索提问词与数据库文献记录中的标引词进行对比，

当提问词与标引词匹配一致时,即为命中记录。由此可见,能否准确地检索出用户所需的信息,关键在于能否准确地选择检索语言。

文献检索语言是一种人工语言,用于各种检索工具的编制和使用,并为检索系统提供一种统一的、作为基准的、用于信息交流的一种符号化或语词化的专用语言。因其使用的场合不同,检索语言也有不同的叫法。在存储文献的过程中用来标引文献的是标引语言,用来索引文献的是索引语言,在检索文献过程中则为检索语言。

检索语言按照表达文献主题概念的词语形式分为分类语言和主题词语言,按照表达文献外部特征的词语形式分为题名(书名、刊名、篇名)、著者姓名、单位名称、出版事项、代码/序号(如专利号、报告号、标准号)等等,作为检索标识直接明了,使用时较为简单。

1. 分类语言

分类语言是以学科体系为基础,用号码作为概念标识,按分类编排的检索语言。在我国比较广泛使用的分类语言是《中国图书馆分类法》。

《中国图书馆分类法》(*Chinese Library Classification*),简称《中图法》(CLC),是我国编制出版的一部具有代表性的大型综合性分类法。《中图法》的编制自1971年至今共出版了五版。《中图法》已普遍应用于各类型的图书馆。国内主要大型书目、检索刊物及各类型机读数据库等都著录《中图法》分类号。

《中图法》的分类体系是根据图书资料的特点,按照从总到分,从一般到具体的编制原则,确定分类体系,在五个基本部类的基础上,组成二十二个大类。《中图法》类目表由以下4个部分组成。

(1) 基本部类:分为马克思主义、列宁主义、毛泽东思想、邓小平理论,哲学,社会科学,自然科学,综合性图书五大类部。

(2) 基本大类:构成分类表的第一级类目。《中图法》为22个基本大类。

表1-1为《中图法》基本部类和基本大类表。

表1-1 《中图法》基本部类和基本大类表

基本部类	基本大类
马克思主义、列宁主义、毛泽东思想、邓小平理论	A 马克思主义、列宁主义、毛泽东思想、邓小平理论
哲学	B 哲学、宗教
社会科学	C 社会科学总论 D 政治、法律 E 军事 F 经济 G 文化、科学、教育、体育 H 语言、文字 I 文学 J 艺术 K 历史、地理
自然科学	N 自然科学总论 O 数理科学和化学 P 天文学、地球科学 Q 生物科学 R 医药、卫生 S 农业科学 T 工业技术 U 交通运输 V 航空、航天 X 环境科学、安全科学
综合性图书	Z 综合性图书

(3) 简表:由基本大类与由其直接展开的一、二类目所形成的类目表。简表反映整个分类表的概况,起着承上启下的作用。

(4) 详表:由简表展开的各种不同登记的类目所组成的类目表,是文献分类的真正依

据。图 1-1 为《中图法》详表摘录（以经济学类为例）。

F	经济
F0	政治经济学
.....	
F1	世界各国经济概况、经济史和经济地理
.....	
F2	经济计划与管理
F20	国民经济管理
F21	经济计划
F22	经济计算、经济数学方法
F23	会计
F239	审计
F24	劳动经济
F25	物资经济
F27	企业经济
F28	基本经济建设
F29	城市与市政经济
F3	农业经济
.....	
F4	工业经济
.....	
F8	财政、金融

图 1-1 《中图法》详表

从图 1-1 可以看出，《中图法》的标记符号采用汉语拼音字母与阿拉伯数字相结合的混合号码，即用一个字母表示一个大类，以字母的顺序反映大类的序列。字母后用数字表示大类以下类目的划分，数字的编号使用小数制。详表中类目的排列反映出学科之间的内在联系。

2. 主题词语言

主题词语言是以语词作为概念标识，按字顺编排的检索语言。主题词语言很多，如关键词语言、单元词语言、标题词语言、叙词语言等，它们有不同的主题词表。主题词表达概念本身，在主题词表中通过参照系统来指示词汇之间的关系。

(1) 关键词语言：关键词（keyword）语言是自然语言，直接取自文献的题名、文摘等，或者取自全文。除了禁用词，如一些冠词、介词、副词或连词外，凡在概念上有意义的词都可用作关键词，它确保检索用词与文献记录中的词汇完全一致。

(2) 单元词语言：单元词语言（uniterm）是规范语言，它是一种最基本的、不能再进一步分割的单位词语言，它能独立表达某一概念。单元词强调单元化词的组配，仅限字面组配。单元词表比较简单，简单的单元词表只有一个字顺表，较完备的单元词表则由一个字顺词表和一个分类词表组成。单元词字顺表包括全部单元词和大量非单元词，非单元词列在单元词条目下，或有参照指向。单元词检索具有灵活、自由组配的特点。

(3) 标题词语言：标题词（heading）语言是一种先组式规范语言，词之间的关系早已有词表规范表达，词表按字顺编排，也有参照指向，实现相关概念的连接。

(4) 叙词语言：叙词（descriptor）语言是一种后组式规范语言，它是为克服单元词与标题词的缺陷而产生的一种较新的、使用较多的检索语言。叙词语言的基本成分是叙词，叙词概念和规范程度都比标题词强，叙词有严格的同义规范、词义规范、词类规范及词形规范。叙词和单元词相同，也有组配性，但它的组配是概念而不是字面组配，两个及两个以上的叙词组合在一起可形成一个新概念。叙词可采用词组形式，以提高其在反映概念上的准确性。叙词的有序排列构成叙词表，它是规范化的、动态性的词表。

1.3.2 信息检索的过程

文献检索是一项实践性很强的活动,它要求我们善于思考,并通过经常性的实践,逐步掌握文献检索的规律,从而迅速、准确地获得所需文献。

信息检索过程通常包括分析检索课题、选择检索工具、确定检索策略、调整检索策略及获取原始文献等步骤。

1. 分析研究课题,明确检索要求

分析检索课题就是明确课题所要解决的问题,即课题所包含的概念和具体要求及相互之间的关系。这是制订检索策略的根本出发点,也是决定检索效率高低的关键。

(1) 分析课题的主题内容:分析课题的主题内容、所属学科性质,明确研究课题所需的信息内容,从而提出能准确反映课题核心内容的主题概念。

(2) 确定课题的文献类型:通过对课题进行主题分析,确定所需信息的文献类型。属于基础理论性探讨的,要侧重于查找期刊论文、学位论文、会议论文;如是尖端技术,应侧重于科技报告;属于发明创造,技术革新,则应侧重于专利文献;如为产品定型设计,则需利用标准文献及产品样本。另外要明确课题对检索深度的要求,即需要提供的是题录、文摘还是原始文献。

(3) 确定检索时间范围:根据课题研究的起始年代和研究的高峰期确定检索的时间范围。

(4) 分析用户的检索评价要求:分析用户的检索评价指标是查新、查准还是查全。一般来说,若要了解某学科、理论、课题、工艺过程等最新进展和动态,则要检索最新的文献信息,重点在于“查新”;若要解决研究中某具体问题,找出技术方案,则要检索有针对性、能解决实际问题的文献信息,重点在于“查准”;若要了解一个全过程、写综述、作鉴定、报成果,就要回溯大量文献,要求检索的全面、详尽、系统,重点在于“查全”。

2. 选择检索工具

(1) 根据检索课题所涉及的学科范围选择检索工具:如果课题复杂,涉及的学科范围广泛,应选择综合性和专业性检索工具相结合的方法进行检索;如果课题涉及的学科领域单一,则应以专业性检索工具为主,综合性检索工具为辅;若检索课题涉及某种特定的信息类型,则应以收录特定信息类型的专业检索工具为主,比如,要检索专利文献,则应重点检索专利数据库,要检索的文献类型是学位论文,则应重点检索学位论文数据库。

(2) 根据检索工具收录范围和编制体系选择检索工具:不同的检索工具在收录范围、报道形式、编排体系、著录格式和索引系统等方面有不同之处,选择检索工具时可根据检索要求加以选择。对于相似的检索工具,可从收录的信息量、文摘和题录信息质量、全文质量、更新速度以及使用的方便程度等方面加以比较。

3. 确定检索策略

(1) 确定检索途径:选用具体的检索工具后,就要确定检索点,以便具体进行检索。各种检索工具提供的检索途径不同,主要分为主题检索(著者、篇名、会议名称、机构等)和分类检索。

(2) 制订检索策略:检索需要制订一种可执行的方案,即检索式。检索式是检索策略