

信息检索教程

主 编 沈艳红

副主编 彭奇志

参 编 严而清 张群 温怀琴



高等教育出版社
HIGHER EDUCATION PRESS

信息检索教程

Xinxi Jiansuo Jiaocheng

主 编 沈艳红
副主编 彭奇志
参 编 严而清 张 群
温怀琴



高等教育出版社·北京
HIGHER EDUCATION PRESS BEIJING

图书在版编目 (C I P) 数据

信息检索教程/沈艳红主编. -- 北京:高等教育出版社, 2012.9

ISBN 978 - 7 - 04 - 036013 - 4

I. ①信… II. ①沈… III. ①情报检索 - 高等学校 - 教材 IV. ① G252.7

中国版本图书馆 CIP 数据核字 (2012) 第 181429 号

策划编辑 柳秀丽
责任校对 杨凤玲

责任编辑 柳秀丽
责任印制 朱学忠

封面设计 于文燕

版式设计 杜微言

出版发行 高等教育出版社
社 址 北京市西城区德外大街 4 号
邮政编码 100120
印 刷 河北省财政厅票证文印中心
开 本 787mm × 1092mm 1/16
印 张 12
字 数 280 千字
购书热线 010 - 58581118

咨询电话 400 - 810 - 0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.landraco.com>
<http://www.landraco.com.cn>
版 次 2012 年 9 月第 1 版
印 次 2012 年 9 月第 1 次印刷
定 价 27.80 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换

版权所有 侵权必究
物 料 号 36013 - 00

前 言

早在 1985 年美国教育界就提出, 面向 21 世纪的学生, 除了要接受传统的阅读、写作和数学教育外, 还需要具有信息交流、批判性思考和解决问题的能力。让每个学生学会如何识别所需信息, 如何检索、组织, 并能以明晰和有说服力的方式描述信息, 是 21 世纪教育最基本的目标之一。

近年来, 信息生态环境变化巨大, 尤其是伴随网络化、数字化而产生的大数量、多类型、多媒体、非规范、跨地域、跨行业、跨语种的网络信息资源, 相对于原来以相对集中和规范为标志的传统文献资源有了突破性发展。在这一全新的信息环境中, 信息资源的组织、查询与利用方式随之发生了根本性变革, 培养学生获取和利用信息的能力日益成为了高等教育教学活动的基本要求。信息检索课程作为大学生信息素质培养和提高的一个重要途径, 其课程内容必须随着信息环境的变化而不断更新。本书便是为适应当前信息环境、实现高等学校人才培养目标编写而成的一部“信息检索”课程教材。

本书内容主要包括以下八个部分: 绪论、信息检索知识、检索工具与参考工具、中文数据库的检索、常用外文全文数据库、特种文献检索、网络信息检索和信息的评价、管理、分析与合理利用。本书在注重培养学生信息检索能力的同时, 也兼顾了对学生在信息评价、管理及其合理利用等能力的培养。此外, 海量的文献资源与人们有限的阅读时间、利用能力之间存在着尖锐矛盾, 这就要求学生具有较强的信息分析能力。为此, 本书特在第 8 章系统介绍了文献计量分析的理论、方法和实际应用。同时在介绍 Web of Science、EI、CNKI、万方等拥有强大的数据挖掘功能的检索系统时, 详细阐述了它们的统计分析功能及其在实际学习、科研中的应用。

全书由沈艳红统筹策划, 各章节的具体编写情况为: 沈艳红编写第 1、2、4 章, 第 3 章第 1 节和第 8 章第 3 节; 严而清编写第 3 章第 2 节; 彭奇志编写第 5、6 章; 温怀琴编写第 7 章; 张群编写第 8 章第 1、2、4 节。最后由沈艳红、彭奇志整理、定稿。

本书是在充分汲取以往此类教材的编写经验和教训, 结合多年来的教学科研成果, 参阅大量信息检索教材的基础上编著而成的, 切实实现了内容精炼、层次清晰、结构合理、实用性强、系统性好的编写目标。本书编写期间得到了一些领导和教师的大力支持, 在此表示衷心的感谢!

由于信息技术的发展日新月异, 加之作者学识和水平有限, 书中难免有疏漏和不足之处, 恳请读者批评指正。

编 者

2012 年 7 月

目 录

第1章 绪论	1	2.4.2 信息检索标识的类型	16
1.1 信息、知识、情报、文献	1	2.4.3 信息检索标识的确定原则	16
1.1.1 信息	1	2.5 计算机检索技术	17
1.1.2 知识	2	2.5.1 布尔逻辑检索	17
1.1.3 情报	2	2.5.2 位置检索	17
1.1.4 文献	3	2.5.3 截词检索	18
1.1.5 信息、符号、文献和载体之间的关系	3	2.5.4 词组检索	18
1.2 信息资源	4	2.5.5 字段检索	19
1.2.1 信息资源的含义	4	2.5.6 限制检索	19
1.2.2 信息资源的类型	4	2.5.7 优先算符	19
1.3 信息素质	6	2.6 信息检索的步骤	19
1.3.1 信息素质的概念	6	2.6.1 分析检索课题	19
1.3.2 信息素质标准	6	2.6.2 选择检索工具或数据库	20
思考题	7	2.6.3 确定检索标识	20
参考文献	7	2.6.4 构造检索表达式	20
第2章 信息检索基础知识	8	2.6.5 确定检索途径	20
2.1 信息检索与效果评价	8	2.6.6 执行检索并评价检索结果	21
2.1.1 信息检索的定义	8	2.6.7 检索策略的调整	21
2.1.2 信息检索的发展历程	8	思考题	23
2.1.3 信息检索的类型	8	参考文献	23
2.1.4 信息检索效果评价	9	第3章 检索工具与参考工具	24
2.2 信息检索语言	10	3.1 检索工具	24
2.2.1 信息检索语言的定义、作用 and 种类	10	3.1.1 检索工具概述	24
2.2.2 分类语言	11	3.1.2 常用检索工具	25
2.2.3 主题语言	13	3.2 参考工具书	36
2.3 信息检索的途径	14	3.2.1 参考工具书概述	36
2.3.1 内容特征途径	15	3.2.2 常用参考工具书举要	40
2.3.2 外表特征途径	15	思考题	59
2.4 信息检索标识	16	参考文献	59
2.4.1 检索标识的定义	16	第4章 中文数据库的检索	61
		4.1 CNKI	61

4.1.1 概述	61	5.3.1 概述	104
4.1.2 检索方法	61	5.3.2 检索技术	105
4.1.3 检索结果处理	67	5.3.3 检索方式	105
4.1.4 CNKI 的其他服务	73	5.3.4 检索结果	106
4.2 万方数据知识服务平台	76	5.3.5 课题检索实例	106
4.2.1 概况	76	5.4 IEL 数据库	106
4.2.2 检索方式	76	5.4.1 概述	106
4.2.3 检索结果处理	79	5.4.2 检索技术	107
4.2.4 知识服务功能	83	5.4.3 检索方式	107
4.3 维普	85	5.4.4 检索结果	109
4.3.1 概况	85	5.4.5 个性化功能	109
4.3.2 检索方式	85	5.4.6 课题检索实例	109
4.3.3 检索结果处理	88	思考题	110
4.4 超星	90	参考文献	110
4.4.1 概况	90	第 6 章 特种文献检索	111
4.4.2 超星数字图书馆	91	6.1 专利与专利检索	111
4.4.3 读秀中文学术搜索	92	6.1.1 专利基础知识	111
4.4.4 百链	94	6.1.2 国际专利分类法	113
4.5 书生之家	95	6.1.3 中国专利数据库	114
4.5.1 概况	95	6.1.4 国外专利数据库	118
4.5.2 检索方式	95	6.2 科技报告	127
思考题	98	6.2.1 概述	127
参考文献	98	6.2.2 我国科技报告	128
第 5 章 常用外文全文数据库	99	6.2.3 国外科技报告	130
5.1 SDOL 数据库	99	6.3 学位论文	132
5.1.1 概述	99	6.3.1 概述	132
5.1.2 检索技术	99	6.3.2 中国学位论文数据库	133
5.1.3 检索方式	100	6.3.3 国外学位论文数据库	134
5.1.4 检索结果	101	6.4 会议文献	138
5.1.5 个性化功能	101	6.4.1 概述	138
5.1.6 课题检索实例	102	6.4.2 中国会议文献数据库	139
5.2 Wiley 数据库	102	6.4.3 国外会议文献数据库	139
5.2.1 概述	102	6.5 标准文献	140
5.2.2 检索方式	103	6.5.1 概述	140
5.2.3 检索结果	103	6.5.2 国内标准文献检索	141
5.2.4 课题检索实例	104	6.5.3 国外标准文献检索	142
5.3 Springer Link 电子期刊	104	思考题	147

参考文献.....	147	第 8 章 信息的评价、管理、分析和 合理利用.....	163
第 7 章 网络信息检索.....	148	8.1 学术信息资源评价.....	163
7.1 Internet 概述.....	148	8.1.1 国内学术信息资源评价体系.....	163
7.1.1 Internet 的起源与发展.....	148	8.1.2 国外学术信息资源评价体系.....	164
7.1.2 Internet 的特点.....	149	8.2 文献管理.....	165
7.1.3 Internet 基本概念.....	149	8.2.1 Note Express.....	165
7.2 搜索引擎.....	150	8.2.2 Endnote.....	169
7.2.1 搜索引擎概述.....	150	8.3 信息分析.....	169
7.2.2 搜索引擎的工作原理.....	150	8.3.1 信息分析概述.....	169
7.2.3 典型的搜索引擎.....	151	8.3.2 文献计量分析.....	170
7.3 开放存取.....	159	8.3.3 信息计量分析应用软件.....	174
7.3.1 开放存取的概念、特征和 意义.....	159	8.4 学术规范与信息合理使用.....	175
7.3.2 开放存取获取途径.....	160	8.4.1 学术规范与学术不端.....	175
7.3.3 常用的开放存取资源.....	160	8.4.2 合理使用.....	176
思考题.....	162	8.4.3 参考文献著录格式.....	176
参考文献.....	162	8.4.4 学术不端文献检测系统.....	178
		思考题.....	179
		参考文献.....	179

第1章 绪 论

信息是一种极其重要的社会财富，它与物质和能量一样，已成为人类社会的重要战略资源。物质提供材料，能量提供动力，信息提供知识。信息化网络化社会的到来，使信息检索从少数检索专家的“专利”转变为大众的基本技能，信息检索已不仅仅局限于科学研究及其相关领域，而是深入到人们工作、学习、生活的方方面面。

1.1 信息、知识、情报、文献

1.1.1 信息

综观人类进化史，它也是一部人类信息活动的发展史。人类社会的发展至今，先后经历了五次信息革命，即语言的产生，文字的出现，印刷术的发明，电报、广播、电视的发明和普及应用，计算机的普及应用及与现代通信技术的结合。第一次信息革命使人类的信息活动从具体跃变为抽象；第二次信息革命则使信息的传递突破了时空的限制；第三次信息革命印刷术使人类记录信息的成本大幅降低；第四次信息革命则在丰富信息存储形式的同时，还大大加快了信息传递的速度；今天仍在持续的第五次信息革命则让人类突破了信息共享的时空限制。

追溯中国文化发展，汉语中的“信息”一词，最早见于《三国志》中“正数欲来，信息甚大”，晚唐诗人杜牧在《寄远》一诗中写道：“塞外音书无信息，道傍车马起尘埃。”这里的“信息”是指音信、消息。信息作为一个科学术语被提出和使用，主要源于西方信息科学的兴起。1928年，R. V. 哈特莱发表了《信息传输》一文，文章认为，信息是指有新内容、新知识的消息。而对信息进行科学定义，则要迟至1948年，C.E. 香农在信息论的奠基性文献——《通信的数学理论》中给出了信息的数学定义和信息量的概念，并提出了信息熵的计算方法，他将“信息”定义为“用以消除随机不确定性的东西”，指出“信息是确定性的增加”。控制论之父 N. 维纳也在其专著《控制论——动物和机器中的通信和控制问题》中对“信息”这一科学概念作出界定，认为信息是“我们在适应和控制外部世界的过程中，同外部世界进行交换的内容的总称”。我国情报学家严怡民在其主编的《情报学概论》一书中指出，信息可以被定义为生物以及具有自动控制系统的机器通过感觉器官和相应的设备与外界进行交换的一切内容。

由于信息内涵丰富、外延宽广，目前尚无一个语意清晰、概念准确的定义。可以这样归纳，所谓信息，是应用文字、信号或者数据等形式描述的客观实在。它表征和维系事物内部结构和外部的联系，感知、表述并反映其属性和差异的状态和方式。同时，信息也是指对诸如事实、数据或观点之类的知识的传递或描述，这些知识可以存在于任何媒体或形式之中，包括文本形式、数字形式、图表形式、图形形式、叙述形式或视听形式。并且，信息将为接受者所接受、理解，更重要的是利用。

从确定性的角度来看，信息具有如下特征：

① 可识别性。信息是可以识别的,识别又可分为直接识别和间接识别,直接识别是指通过感官的识别,间接识别是指通过各种测试手段的识别。不同的信息源有不同的识别方法。

② 可存储性。信息是可以通过各种方法存储的。

③ 可扩充性。信息随着时间的变化,将不断扩充。

④ 可压缩性。人们对信息进行加工、整理、概括、归纳就可使之精练、浓缩。

⑤ 可传递性。信息可以通过传播媒体进行传递、传播。

⑥ 可转换性。信息可以由一种形态转换成另一种形态。

⑦ 特定范围有效性。信息在特定的范围内是有效的,否则是无效的。

1.1.2 知识

知识是人类哲学研究的重要范畴。在《中国大百科全书》(哲学卷)中指出,知识是人类认识的成果,是在实践的基础上产生又经过实践检验的对客观实际的反映,就它的反映活动形式而言,有时表现为主体对事物的感性知觉或表象,属于感性知识,有时表现为关于事物的概念或规律,属于理性知识。《辞海》(1999年缩印版)对知识的解释为人类认识的成果或结晶。依反映对象的深刻性,可分为生活常识和科学知识;依反映层次的系统性,可分为经验知识和理论知识。

世界经济合作与发展组织(Organisation for Economic Co-operation and Development, OECD)在1996年的年度报告《以知识为基础的经济》中将知识分为四大类:

① 知道是什么的知识(Know - What)。主要是叙述事实方面的知识。

② 知道为什么的知识(Know - Why)。主要是自然原理和规律方面的知识。

③ 知道怎么做的知识(Know - How)。主要是指对某些事物的技能和能力。

④ 知道是谁的知识(Know - Who)。涉及谁知道和谁知道如何做某些事的知识。

德国科学家、知识管理学家迈克尔·波兰尼认为,“人类的知识有两种。通常被描述为知识的,即以书面文字、图表和数学公式加以表述的,只是一种类型的知识。而未被表述的知识,像我们在做某事的行动中所拥有的知识,是另一种知识。”他将前者称为显性知识,而将后者称为隐性知识。

知识具有以下主要属性:

① 实践性。实践是知识产生的基础,也是验证知识的标准,正确的知识又可以指导人们的实践活动。

② 规律性。知识揭示了事物及其运动过程的规律性。

③ 渗透性。随着知识门类的增多,各学科知识之间相互交叉、渗透,出现了很多新的知识门类。

④ 继承性。一方面,新知识的产生,既是原有知识的深化与发展,又是更新的知识产生的基础和前提;另一方面,通过各种方式被记录下来的知识可以世代相传。

1.1.3 情报

人们最早把情报当作战时有关敌情的报告。19世纪到20世纪50年代,情报被视为是一种系列化的知识。20世纪70年代,情报被认为是人们决策、部署、规划和行为、意志所需要的

能指引方向的知识与智慧。目前,情报被认为“是被有目的利用的、活化的知识与信息”。按照应用范围,情报可分为科学情报、经济情报、技术经济情报、军事情报、政治情报等;按内容及作用分类,情报则可被划分为战略性情报和战术性情报两大类。

目前常用的情报研究方法主要包括社会调查法、文献计量统计方法、数学分析法、系统分析与评价方法、历史的研究方法等。其中与信息检索课程最为相关的莫过于文献计量统计方法,它是对以记录形式进行交流的各个方面进行计量统计,如对于某国、某学科的论文数量、年度增长率、利用状况、对文献中所附参考文献种类、数量等进行计量统计,从中找出变化规律,建立相应数学模型,从定性定量分析中达到掌握过去与现在的变化脉络,进而预测未来可能的变化,在本书第8章对此做了简单介绍。

1.1.4 文献

文献一词在我国最早见于《论语·八佾》。宋代朱熹解释为“文,典籍也;献,贤也。”即记载和掌握知识的书籍和有学问的人是“文”和“献”。《中华人民共和国国家标准·文献著录总则》(GB3792.1—83)赋予文献的定义是“记录有知识的一切载体”。

文献有四个基本构成要素:①信息内容,即文献所表达的思想意识、知识信息的涵义和内容;②符号系统,是指揭示和表达知识、信息内容的标识符号,是物化和标识文献信息内容的工具,如文字、图形、编码、声频、视频等;③文献载体,是承载文献信息符号,从而使信息内容有所依附并便于传播交流的物质材料,一般可分为纸质和非纸质两大类,非纸质载体材料很多,如甲骨、金石、贝叶、兽皮、胶片、磁带、磁盘、光盘等;④记录方式,是指将包含信息内容的信息符号存储到载体材料上的方式,如书写、雕刻、印刷、拍摄、录制、复印、刻录等。

1.1.5 信息、符号、文献和载体之间的关系

知识、文献、情报之间的关系可用维恩图来表示。图1.1.1中A表示信息,B表示知识,C表示文献,用D来表示情报。信息中的一部分被人们所感知、所认识,并在反复的认识—实践—再认识—再实践的过程中,由感性认识上升到理性认识,这就形成了知识;部分信息与知识被载体所记录方形成文献;为特定目的而有意识地搜集的,有使用价值的文献、信息与知识为情报。

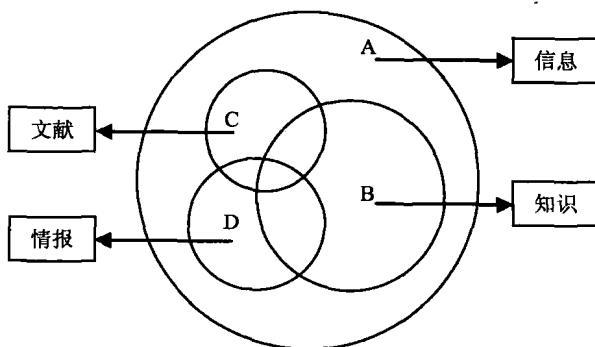


图 1.1.1 信息、知识、情报、文献关系图

1.2 信息资源

1.2.1 信息资源的含义

与信息一样,目前对信息资源的认识和理解也未能达成共识。1979年,美国管理学者霍顿对信息资源的解释是:单数形式的信息资源(Information Resource)是指某种内容的来源,即包含在文件和公文中的信息内容;复数形式的信息资源是指支持工具,包括供给、设备、环境、人员和资金等。中国学者乌家培教授认为,狭义的信息资源仅指信息内容本身;广义的信息资源既包括信息内容,还包括与其紧密相联的信息设备、信息人员、信息系统、信息网络等。

信息资源是影响社会发展和进步的一种重要战略资源,多数专家认同从狭义和广义两种角度来认识和理解其涵义。狭义的信息资源是指人类社会活动中经过加工处理的、有序化的、并大量累积后的有用信息的集合。广义的信息资源则是指信息和它的生产者以及信息技术的集合,一般由三部分构成:

- ① 人类社会活动中经过加工处理有序化并大量累积后的有用信息的集合。
- ② 为某种目的而生产有用信息的信息生产者的结合。
- ③ 加工、处理和传递有用信息的信息技术的集合。

1.2.2 信息资源的类型

信息资源种类繁多,数量庞大,可以按照不同的标准对其进行分类。目前通行的有三种分类方法,分别是按加工程度划分、按载体类型划分和按存在状态划分。

1. 按加工程度划分

根据信息加工程度的不同,可以将信息资源划分为零次信息资源、一次信息资源、二次信息资源和三次信息资源。

零次信息资源指的是未向社会公开和正式报道,只提供一定范围内使用的信息资源,主要包括形成一次文献以前的知识信息,是未经记录、未形成文字材料、未公开于社会,即未经正式发表的原始文献,如书信、手稿、笔记、设计草图、实验记录、会议记录、网络聊天记录等,一般通过口头交谈、参观展览、参加报告会等途径获取。

一次信息资源是指人们根据自己的科学实验、生产实践的成果而撰写,未经他人加工、重组、压缩,通过某种方式向社会公开发布或报道的信息资源,主要包括专著、教材、学术期刊论文、科技报告、会议论文、学位论文、专利说明书等。

二次信息资源指将各种分散、无序的一次信息资源,按照一定的规则整理、归纳、提炼和简化,按一定的规律组织而形成的供查找一次信息资源的线索型信息资源,因此常称其为检索工具。二次信息资源主要包括目录、题录、文摘和索引等类型,可以帮助信息检索人员有效地提高检索和使用一次信息资源的效率。

三次信息资源是在对一次信息资源和二次信息资源进行评价和筛选后,经过分析、提炼和压缩,按知识门类或专题综合加工而成的信息资源,所以我们往往称它们为参考工具。三次信息资源主要包括百科全书、年鉴、字词典、手册等,具有概括性、浓缩性、参考性和指导性。

概括地说,零次信息资源往往是一次信息资源的素材;一次信息资源是信息资源的基本形

式,是信息检索人员检索的主要对象;二次信息资源是检索一次信息资源的工具;三次信息资源则是对一次信息资源和二次信息资源的综合分析研究结果。

2. 按载体的类型划分

按照信息载体类型的不同,信息资源又可被划分为纸质资源和非纸质资源两大类。

纸质信息资源是以纸张作为载体的信息资源,按照其记录方式的不同,又可以细分为印刷型和书写型两类。印刷型信息资源包括图书、期刊、报纸等;书写型信息资源主要包括手稿、笔记等。

非纸质信息资源是指以纸张以外介质作为载体的信息资源,除了古代的甲骨、竹简、贝叶等之外,现代的非纸质信息资源主要包括以下三类:①缩微型信息资源:是以感光材料为载体,以缩微照相技术为记录手段的一种比纸质信息资源缩小若干倍的信息资源,主要包括缩微胶卷、缩微平片等。②声像型信息资源(又称视听型信息资源):是以磁性材料和感光材料为存储介质直接记录声像信息的信息资源,包括唱片、磁带、录音带、录像带、幻灯片等。③电子型信息资源:是指以二进制的格式把文字、声音、图像和动画等存储在磁盘、光盘等载体上,需要通过计算机、网络通信或终端等设备进行阅读的信息资源,包括光盘数据库、网络数据库、联机数据库、电子图书、电子期刊、网络信息资源等。

3. 按存在状态划分

按信息的存在状态,信息资源还可分为潜在信息资源和现实信息资源。潜在信息资源是指个人在认知和创造过程中储存在大脑中的信息资源,它们虽能为个人所利用,但一方面易于随忘却过程而消失,另一方面又无法为他人直接利用。现实信息资源是指潜在信息资源经个人表述之后能为他人所用的信息资源,主要包括人体载体信息资源、文献信息资源、实物信息资源和网络信息资源。

① 人体载体信息资源。按其表述方式又包括口语信息资源和体语信息资源。口语信息资源是人类以口头语言所表述出来的信息资源,如谈话、授课、讲演、讨论、歌唱等。体语信息资源是以人的体态表述出来的信息资源,如表情、手势、姿态、舞蹈等。

② 实物信息资源。是指以实物为载体的信息资源。依据实物的人工与天然特性,又可将实物信息资源分为以自然物质为载体的天然实物信息资源(如土壤、岩石、植被等)和以人工实物为载体的人工实物信息资源(如样品、模型、雕塑等)。

③ 文献信息资源。是指以文字、图形、符号、声频、视频等方式记录在各种载体上的信息资源,也是目前内容最丰富、使用频率最高的信息源。国家标准《文献类型与文献载体代码》(GB/T 3469—1983)根据实用标准,将文献划分为26个类型,即专著、报纸、期刊、会议录、汇编、学位论文、科技报告、技术标准、专利文献、产品样本、中译本、手稿、参考工具、检索工具、档案、图表、古籍、乐谱、缩微胶卷、缩微平片、录音带、唱片、录像带、电影片、幻灯片、其他(盲文等)。

④ 网络信息资源。是指计算机技术、通信技术、多媒体技术相互融合而形成的网络上可查找到的信息资源,按信息表现形式分有电子出版物和非电子出版物两类。电子出版物是指以电子方式或机读方式生产和发行的出版物。电子出版物的文字、图像和声音信息是以数字方式存储在磁盘和光盘等介质上的,通过电子计算机输出设备和网络在视频终端上显示出来。目前出版的各种电子出版物有电子期刊、电子报纸、电子图书、电子名录、电子地图、数字声音(如

CD)和数字图像(如VCD、DVD)等。其中,全文数据库、二次文献数据库和新一代联机公共检索目录(Online Public Access Catalog, OPAC)、数值数据库、事实数据库是最受用户欢迎、最具代表性的电子出版物。非电子出版物信息资源主要是指电子新闻、电子论坛等。

1.3 信息素质

1.3.1 信息素质的概念

信息素质一词,最早是在1974年由美国信息工业协会的会长Paul Zurkowski首次提出的。当时他对信息素质下的定义是:利用大量的信息工具及主要信息源使问题得到解答的技术和技能。1979年,全美信息产业协会给出了信息素质的定义是:掌握了利用信息工具的知识与技能,并将其应用于解决实际问题。1989年,美国图书馆协会(American Library Association, ALA)发表的信息素质研究报告指出,具备信息素质的人,是能够敏锐地察觉信息需求,并能够进行相应的信息检索、评估以及有效利用所需信息的人。

1.3.2 信息素质标准

1. 美国图书馆协会和教育传播与技术协会制定的信息素质标准

1998年,美国图书馆协会和教育传播与技术协会制定了学生学习的九大信息素质标准,这一标准从信息素质、独立学习和社区责任三个方面进行表述,丰富了信息素质的内涵,具体内容如下:

(1) 信息素质

标准一:具有信息素质的人能高效获取信息。

标准二:具有信息素质的人能熟练地、批判性地评价信息。

标准三:具有信息素质的人能够精确地、创造性地使用信息。

(2) 独立学习

标准四:作为一个独立的学习者,具有信息素质,并能探求与个人兴趣有关的信息。

标准五:作为一个独立的学习者,具有信息素质,并能欣赏作品和其他对信息进行创造性表达的内容。

标准六:作为一个独立的学习者,具有信息素质,并能在信息查询和知识创新中力争做到最好。

(3) 社会责任

标准七:对学习社区和社会有积极贡献的人,具有信息素质,并认同信息对社会的重要性。

标准八:对学习社区和社会有积极贡献的人,具有信息素质,并能实行与信息 and 信息技术相关的符合伦理道德的行为。

标准九:对学习社区和社会有积极贡献的人,具有信息素质,并能积极参与活动来探求和创建信息。

2. 我国的信息素质标准

我国在1999年《中共中央国务院关于深化教育改革全面推进素质教育的决定》中规定:“在

高中阶段的学校和有条件的初中、小学普及计算机操作和信息技术教育”，并提出了培养学生信息素养的六个标准：

- ① 对信息的关注。
- ② 能够研究和判别可供选择的信息及观点的优劣、可行性。
- ③ 能够选择各种信息源，熟悉使用各种信息工具以获取信息的能力。
- ④ 能对某一课题找到相应的信息，并进行综述及纵观其现状和发展趋势。
- ⑤ 获得继续自我教育的基础。
- ⑥ 发挥主动性和独立性。

2000 年，我国《中小学信息技术课程指导纲要（试行）》提出了六个方面的信息素养教育和培养目标：

- ① 信息获取能力。
- ② 信息分析能力。
- ③ 信息加工能力。
- ④ 信息创新能力。
- ⑤ 信息利用能力。
- ⑥ 信息意识和信息交流的能力。

清华大学孙平教授主持研制了《北京地区高校信息素质能力指标体系》，并于 2005 年向社会发布，该指标体系由 7 个一级指标、19 个二级指标、61 个三级指标组成。7 个一级指标为：

- ① 能够了解信息以及信息素质能力在现代社会中的作用、价值与力量。
- ② 能够确定所需信息的性质与范围。
- ③ 能够有效地获取所需要的信息。
- ④ 能够正确地评价信息及其信息源，并且把选择的信息融入自身的知识体系中，重构新的知识体系。
- ⑤ 能够有效地管理、组织与交流信息。
- ⑥ 作为个人或群体的一员，能够有效地利用信息来完成一项具体的任务。
- ⑦ 了解与信息检索、利用相关的法律、伦理和社会经济问题，能够合理、合法地检索和利用信息。

思考题

1. 试述信息、知识、情报、文献的定义以及四者之间的关系。
2. 按加工程度划分，信息资源可分为哪些类型？
3. 试述信息素质的定义。

参考文献

- [1] 彭奇志. 信息检索与利用教程[M]. 北京：中国轻工业出版社，2009.
- [2] 喻萍. 现代经济信息检索与利用[M]. 北京：化学工业出版社，2010.
- [3] 许征尼. 信息素养与信息检索[M]. 合肥：中国科学技术大学出版社，2010.

第2章 信息检索基础知识

2.1 信息检索与效果评价

2.1.1 信息检索的定义

广义的信息检索（Information Retrieval）是指将分散、无序的信息按照一定的方式组织和存储起来，并根据用户的信息需求，依据相应的规则从信息集合中找出有关信息的过程，因此又称为信息的存储与检索。而通常人们只把信息检索理解为整个过程的后半部分，即信息查找的过程。

2.1.2 信息检索的发展历程

信息检索起源于 19 世纪前期。当时，文献量逐渐增多，需要有人能对所有发表的文献及时地进行收集、加工和整理，并提供一定的手段，方便人们查找文献，信息检索工作便由此产生了。19 世纪末 20 世纪初，出现了覆盖各种专业领域的检索工具，其中有一些逐步成为世界闻名的检索工具，如美国的《工程索引》和《化学文摘》、英国的《科学文摘》等。到了 20 世纪四五十年代，出现了一些半机械化、机械化的检索操作方式，例如各种穿孔卡片检索工具。1951 年，世界上出现了最早的计算机检索系统，20 世纪 60 年代，计算机性能的提高、存储量的扩大、价格的下降为计算机信息检索的发展提供了坚实的物质基础。20 世纪 70 年代，随着卫星技术和通信技术的发展，联机信息检索突破了地域的限制，如 DIALOG 等开始向全世界提供商业化的联机信息检索服务。与此同时，许多手工信息检索工具也针对新的社会需求和信息技术的背景，进行了相当程度的调整，进一步趋于成熟。1987 年，DIALOG 系统率先推出光盘数据库检索服务。20 世纪 90 年代，网络检索、多媒体检索以惊人的速度迅速崛起，1994 年出现了第一个 Web 搜索引擎。到 1998 年，网上的搜索引擎数量已达 400 个之多。网络检索以极低的费用、海量的信息、迅速的存取以及对多媒体功能的支持，对联机信息检索、光盘信息检索造成了强大的冲击，成为目前最为常用的检索方式。

2.1.3 信息检索的类型

1. 根据检索对象区分

根据检索对象的不同，信息检索可以区分为文献检索、事实检索和数据检索。

① 数据检索。以数据作为检索对象，查找用户所需要的数值型数据。其检索对象包括各种调查数据、统计数据、特性数据等。例如：查找某一企业的年销售额、某一国家的人口数、物质的属性数据等。

② 事实检索。以事实作为检索对象，例如，查找某一机构、企业或人物的基本情况、查找某一人物的生平等。

③ 文献检索。以文献作为检索对象，查找与用户需求相关的文献信息。例如，查找有关“信息组织”、“无线传感网”的文献等。

对于数据检索或事实检索而言，由于用户所需的数据或事实往往不能脱离文献而单独存在，因此数据或事实检索也是以文献为依托的，它们与文献检索的差异主要在于：对于数据或事实检索，检索人员只需把有关的数据或事实提供给用户，就能直接满足用户的信息需求；而对于文献检索，检索人员通常是把查找出来的相关文献提供给用户，由用户阅读、理解、吸收和利用文献的信息内容、从而满足其信息需求。

2. 按检索方式区分

按检索方式的不同，信息检索可以区分为手工信息检索和机器信息检索。

① 手工信息检索。以手工操作的方式，利用检索工具书进行信息检索。其优点是便于控制检索的准确性，缺点是检索速度慢、工作量较大。

② 机器信息检索。以机械、机电或电子化的方式，利用检索系统进行信息检索。机器信息检索是从 20 世纪 40 年代以后逐渐发展起来，电子计算机诞生之后，以强大的存储能力、不断提高的处理性能以及同步降低的价格，很快便成为机器信息检索的主流和代表。因此，机器信息检索主要就是指计算机信息检索，其优点是检索速度快、能够实现多字段组合检索；相对手工信息检索而言，其缺点主要是需要借助相应的设备进行检索。

3. 按检索的信息类型区分

按检索的信息类型的不同，信息检索可以区分为文本检索和多媒体检索。

① 文本检索。查找含有特定信息的文本文献的检索，其结果是以文本形式反映特定信息的文献。文本检索是一种传统的信息检索类型，在信息检索中占据着主要地位。

② 多媒体检索。查找含有特定信息的多媒体信息的检索，其结果是以多媒体形式反映特定信息，如图像、声音、动画、影片等，是在网络环境下发展起来的全新检索类型。

2.1.4 信息检索效果评价

这里所讨论的信息检索效果的评价，主要是从信息需求者的角度而非信息检索系统的角度出发。对信息需求者来说，自然希望在最短的时间内获取最满意的检索结果，但在实际检索时，往往受诸多因素的影响，只能检索出部分相关信息，同时也会发现检索结果中部分信息其实与自己的需求并不相关。信息检索过程中使用的检索语言、检索方法、检索技术、检索途径等都会影响检索结果，检索者对检索效率进行合理的评价和研究可以提高自己的检索效率。

表 2.1.1 检索结果数量表示

项目	相关文献量	不相关文献数量	合计
检出文献数量	a	b	$a+b$
未检出文献数量	c	d	$c+d$
合计	$a+c$	$b+d$	$a+b+c+d$

常用的检索效果评价指标包括查全率、查准率、误检率、漏检率等。我们将检索结果数量用表 2.1.1 表示，上述指标可分别定义为：

① 查全率。即检出的相关文献与全部相关文献的百分比，其计算公式如下：

$$\text{查全率} = a/a+c$$

② 漏检率。即未检出的相关文献与全部相关文献的百分比，其计算公式如下：

$$\text{漏检率} = c/a+c$$

并且

$$\text{漏检率} = 1 - \text{查全率}$$

③ 查准率。即检出的相关文献与检出的全部文献的百分比，其计算公式如下：

$$\text{查准率} = a/a+b$$

④ 误检率。即检出的非相关文献与检出的全部文献的百分比，其计算公式如下：

$$\text{查准率} = b/a+b$$

并且

$$\text{误检率} = 1 - \text{查准率}$$

对于用户来说，查全率和查准率自然都是越高越好，但是，两者同时达到 100% 几乎是不可能的。在实际的查询过程中，可以发现，当我们想方设法提高查全率的时候，往往会放宽检索范围以检出更多的相关信息，导致部分相关度较低甚至不相关的信息也被误检了出来；反之，当我们想方设法提高查准率的时候，就会使用比较严苛的检索条件以检出的高相关度信息，导致部分相关度略低的信息被漏检。查全率和查准率之间往往存在此消彼长的互逆关系，因此，用户在检索之前，先要衡量对于该次检索查全率和查准率哪个更为重要，在制定检索策略的时候就可以有所侧重。

2.2 信息检索语言

2.2.1 信息检索语言的定义、作用和种类

1. 信息检索语言的定义

信息检索语言是用来描述文献内部特征和外表特征、表达信息提问的一种人工语言，又称标引语言、索引语言、概念标识系统等，是信息检索系统存储和检索信息时共同使用的一种约定性语言。对于信息加工人员来说，信息检索语言是表达信息主题内容、形成信息检索标识并借以组织信息的依据；对于信息检索人员来说，它是表达检索课题要求，并同检索系统中已经存储的信息标识进行比较从而获得所需信息的依据，因此，信息检索语言也是信息加工人员与信息检索人员沟通的桥梁。

2. 信息检索语言的特点

与其他语言相比，信息检索语言突出的特点是：

① 具有必要的语义和语法规则，能准确地表达科学技术领域中任何标引和提问的中心内容和主题；

② 具有表达概念的唯一性，即同一概念不允许有多种表达方式，不能模棱两可；

③ 具有检索标志和提问特征进行比较和识别的方便性；

④ 既适用于手工检索系统又适用于计算机检索系统。

3. 信息检索语言的作用

信息检索语言在信息检索中的作用主要是：