

文献信息检索

教程

WENXIAN XINXI JIANSUO JIAOCHENG

主 编 宋衍武

副主编 蒋 红 李武伟 贾志广

青海人民出版社

文献信息检索教程

主 编 宋衍斌
副主编 蒋 红 李王伟
贾志广

青海人民出版社

· 西 宁 ·

图书在版编目(CIP)数据

文献信息检索教程/宋衍斌主编. —西宁:青海人民出版社, 2006. 9

ISBN 7-225-02825-1

I. 文... I. 宋... II. 情报检索—高等学校—教材 N. G252.7

中国版本图书馆 CIP 数据核字(2006)第 113989 号

文献信息检索教程

WENXIAN XINXI JIANSUO JIAOCHENG

宋衍斌 主编

出版 青海人民出版社 (西宁市同仁路10号)
发行 : 邮政编码 810001 总编室 (0971)6143426
发行部(0971)6143516 6123221
印刷: 青海新华印刷厂
经销: 新华书店
开本: 850mm×1168mm 1/32
印张: 10.5
字数: 270 千
版次: 2006 年 8 月第 1 版
印次: 2006 年 8 月第 1 次印刷
印数: 1—1 000 册
书号: ISBN 7-225-02825-1/G · 1149
定价: 26.00 元

版权所有 翻印必究

(书中如有缺页、错页及倒装请与工厂联系)

· 前 言

呈现在读者面前的这本书，是1998年出版的《图书馆应用教程（现代图书馆）》的姐妹篇。两者神应契合，貌却相异。说是“貌异”是因为这两部书的时间跨度和叙述体例各不相同；说其“神合”是因为这两部书有着相同的企盼，即为我们直接身处其内并使我们应接不暇的图书馆文献信息急剧变动过程提供一点真切的感受与切实的理解。

观察转型过程的图书馆，可以有两个不同的“视点”，每一个视点，可以有两个不同的“视角”。第一个“视点”的两个“视角”是“从外向内看”与“从上往下看”；第二个“视点”的两个“视角”是“从内向外看”与“从下往上看”。何谓“外、内、上、下”？

所谓的“外”，就是西方社会图书馆信息理论与范畴。“由外向内看”就是通过“译语”来考察我们的图书馆信息理论的现代化进程；所谓“内”，即我们自身的历史与现实尤其是依然活跃在人们头脑中的习惯观念和行为方式对我们运用文献信息的影响和作用；所谓“上”，是指整齐划一的文献信息系统。“从上往下看”就是试图通过流行的“官语”来考察我们图书馆信息现代化的过程；所谓“下”，是指与图书馆信息领域相对应的社会领域，尤其是指广大的求知人群。

这两个视点具有相同的重要性。虽然由于“视点”不同，但可相辅相成，共同深化我们对自身所处的信息化社会的认识。

“梦断美人沉信息，目穿长路倚楼台。”南唐李中的一首《暮春怀故人》，使我们感受到了古人对信息的渴望。李清照的

那句“不乞隋珠与和璧，只乞乡关新信息”则让我们体味到了信息在人类生活中的弥足珍贵。正像人类在上个世纪所畅想的那样，全球信息化浪潮，一浪高过一浪。文献信息的数字化，数字信息的网络化，网络信息的海量激增和信息的有效组织、存储和检索，将是我们关注的前沿课题。对知识的收集和管理意义已经变得和对知识的探索创造一样重要，信息已经和物质、能源并列为人类社会发展的三大资源。

人文与科技信息资源，作为人类生存方式的内容和载体，从根本上影响和改变人类的进程。不论是作为文献之邦的中国，还是科技发达的美英等国，都把人文信息资源的占有和利用看作立国富民的头等大事。人类越来越清楚地认识到，在竞争日趋激烈的社会环境中，信息意味着机遇和发展，谁最先获得了有价值的信息，谁就会在竞争中占据优势。

因此，提高当代大学生的人文素质，特别是信息素质已经成为国内外教育界的共识。在这方面，一些发达国家已经走在了我们的前头，从20世纪80年代末，我国开始在高等院校增设文献检索课，以使大学生适应信息时代的需要，尽快掌握信息检索的基本原理与方法，了解信息检索在人文研究过程中的作用和意义，培养人文专业学生的信息意识及计算机和互联网检索技术，使其能够在激烈的竞争中生存和发展。每个大学都有其特定的学科结构、发展历程、建设规模和与之相对应的藏书体系。因此，尽管我们处于一个网络化的信息共享的社会，但各自的信息环境还是存在着一定的差异。为此，我们结合青海师范大学的文献信息资源，为文史、政法、经济、艺术、外语、理工类专业的学生量身定制了这本《文献信息检索教程》。

本书在编写过程中注重教材的针对性、可行性和指导性。力争做到检索理论与检索方法有机结合，手工检索与计算机检索有机结合，信息检索与信息利用有机结合，使教材更具有科学性、系统性和实用性。强调了基本技能、自学能力和研究能力的培养。

在编写本书的过程中，我们更加真切地感受到了信息、情报和文献对我们人类的重要性。对 Google 的创始人 Larry Page 所说的那句话“我们若能更妥善地搜寻资料，实在已经改变了世界”，感同身受。

本书各章执笔者如下：宋衍斌：前言、第二章、第三章、第四章、第五章、第九章、第十章；蒋红：第六章、第七章、第十二章、第十五章；李武伟：第一章、第八章、第十一章、第十四章；贾志广：第十三章、第十六章、第十七章、第十八章。本书最后由宋衍斌统改定稿。

在本书编写过程中，我们参阅了大量中外文资料，在此我们向引文作者表示敬意。由于水平所限，书中不当之处敬请同行和读者斧正。

宋衍斌

2006 年 1 月

目 录

前 言

第一章 信息与信息检索	(1)
第一节 信息及相关概念	(1)
第二节 信息检索	(5)
第三节 计算机信息检索	(12)
第四节 知识与知识管理	(19)
第二章 人文工具书类型及其概述	(22)
第一节 人文文献工具书的性质和作用	(22)
第二节 常用人文文献工具书的类型	(24)
第三节 人文社会科学工具书的主要排检方法	(29)
第四节 数字化人文社会科学工具书	(36)
第三章 字词检索	(38)
第一节 中文字词的检索	(38)
第二节 外文字词的检索	(48)
第四章 书目与古今图书查找	(54)
第一节 古籍的查找	(54)
第二节 近代图书的查找	(59)
第三节 禁书、伪书的查找	(61)
第五章 哲学政治学信息检索	(65)
第一节 马列文献的检索	(65)
第二节 哲学文献的检索	(71)
第三节 政治学信息检索	(75)
第六章 法律法规类文献检索	(78)

第一节	中国法律法规信息的检索工具	(78)
第二节	外国法学文献信息的检索	(83)
第三节	国际条约的检索工具	(85)
第四节	法律法规网络信息资源检索	(87)
第七章	经济与管理学信息检索	(95)
第一节	经济与管理学文献概况及分布特点	(95)
第二节	经济与管理学检索工具及利用	(97)
第三节	统计信息的查找	(102)
第八章	文学信息资源检索	(108)
第一节	文学专题书目检索	(108)
第二节	文学词典检索	(113)
第三节	文学资料汇编检索	(115)
第九章	历史地理信息检索	(119)
第一节	历史信息的检索	(119)
第二节	地理信息的查找	(131)
第十章	特种文献检索	(136)
第一节	会议文献检索	(136)
第二节	政府出版物检索	(142)
第三节	学位论文检索	(145)
第十一章	数据库检索	(150)
第一节	数据库及数据库检索	(150)
第二节	商业数据库检索	(154)
第三节	中文数据库检索	(174)
第十二章	Internet 信息检索	(186)
第一节	Internet 基础知识	(186)
第二节	Internet 主要服务及使用	(190)
第三节	搜索引擎	(195)
第四节	Internet 信息的收集、选择、开发和利用	(201)
第十三章	Internet 信息查询技巧	(205)
第一节	Internet 检索策略	(205)

第二节	Internet 上的电子报纸与新闻	(208)
第三节	Internet 上的工具书	(215)
第四节	Internet 专题信息的查找	(221)
第十四章	图书馆馆藏文献检索	(227)
第一节	认识图书馆	(227)
第二节	ILAS 与图书馆馆藏文献检索	(229)
第三节	TPI 与虚拟馆藏检章	(235)
第十五章	社会科学研究	(238)
第一节	社会科学研究	(238)
第二节	社会科学研究选题	(240)
第三节	社会科学论文的撰写	(248)
第十六章	题录型文献检索工具	(251)
第一节	《全国报刊索引》(科技版)	(251)
第二节	《中文科技资料目录》	(254)
第三节	《国外科技资料目录》	(260)
第四节	美国《近期期刊目次》	(261)
第十七章	文摘型文献检索工具	(272)
第一节	美国《工程索引》	(272)
第二节	英国《科学文摘》	(278)
第三节	美国《化学文摘》	(282)
第四节	日本《科学技术文献速报》	(294)
第五节	《美国生物学文摘》	(296)
第十八章	科学论文引文索引	(303)
第一节	概述	(303)
第二节	中国科学引文索引	(305)
第三节	美国《科学引文索引》	(306)
第四节	ISI Web of Science	(312)
第五节	《默克索引》	(313)
第六节	网络虚拟参考工具书	(318)
参考文献		(322)

第一章

信息与信息检索

第一节 信息及相关概念

我们所处的时代是一个在前所未有的层次上不断产生、传递和利用信息情报的时代。因此，有人称我们的时代为信息时代。那么，究竟什么是信息？什么是情报？它们与本课程大量讨论的文献课题有着什么关系？这是我们首先需要了解的问题。

一、信息

信息——物质存在的一种方式，一般指数据、消息中所包含的意义。可以使消息中所描述的事件的不确定性减少。

不同的学科，从不同的角度对信息这个概念有不同的解释。经济学、心理学、新闻学和哲学的定义不同于图书情报学对信息的定义。控制论专家 N·维纳从信息自身具有的内容属性给信息所下的定义被许多研究者所引用。国内在这方面也有许多争论。

信息的定义之所以呈现多样化，主要原因有三：第一，信息本身的复杂性，它是一个多元化、多层次、多功能的综合物；第二，信息科学是一门新兴学科，它的许多分支学科仍在随着社会、经济和科学技术的发展而发展，其内涵和外延不很确切；第三，人们出于不同的研究和使用目的，从不同的角度或层次出发，对信息概念就会作出不同的解释。

因此，许多学者建议将信息的概念分为不同的层次来解释。在诸多层次中，最重要的是两个层次：一是没有任何约束条件的

本体论层次，即信息是一种客观存在的现象，是事物的运动状态及其变化方式的表征，不受主体意志的影响。不停运动着的事物不断产生本体论意义上的信息；二是受主体约束的认识论层次，即信息就是主体所感知或所表述的事物运动状态及其变化方式，是反映出来的客观事物的属性。

信息是广泛存在的，世间万物的运动，人间万象的更迭，都离不开信息的作用。李太白的诗“日照香炉生紫烟，遥看瀑布挂前川，飞直下三千尺，疑是银河落九天。”给我们带来的显然就是庐山瀑布的信息；苏东坡的词“大江东去，浪淘尽，千古风流人物……”给我们传递的是赤壁怀古的信息。

从自然界角度看，信息表征了物质的属性，地球昼夜的变化是一种信息，它反映出地球绕太阳自转的运动特性和状态；山的高度是一种信息，它反映出山的空间特性；树干的年轮是一种信息，它反映了树木成长的时间特性——树龄；闪电是一种信息，它反映了云层中所含能量的特性；花的香味也是一种信息，它反映了花分子结构的化学特性等。

由此我们可以看出信息并非是区别事物的本质属性，信息并非事物本身，而是事物发出的体现它本质属性的信号。

信息具有如下特征：①客观性；②动态性；③相对性；④依存性；⑤可传递、亦可干扰性；⑥可加工性；⑦共享性。

二、知识

知识——是人类的主观世界对客观世界的概括和反映，是大量有组织的信息，是关于事实和思想的有组织的陈述，它提供某种经过思考、判断和某种实验的结果。

知识是人类社会实践经验的总结，知识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，是人的大脑通过思维重新组合的系统化的信息的集合。因此，人类不仅要通过信息感知世界、认识和改造世界，而且要根据所获得的信息组成知识。可见，知识是信息的一部分。

1. 知识的分类

经济合作与发展组织 (OCED) 出版的《以知识为基础的经济》中将知识分为以下几类:

第一类“知事 (Know - what)”, 指关于事实方面的知识, 也可理解为 Know - when、Know - where; 即在什么样的时间 (Know - when)、什么样的地点或条件下 (Know - where) 能解决什么样的问题。

第二类“知因 (Know - why)”, 指自然原理和规律方面的科学理论, 知识的生产是在专门研究机构如实验室和大学完成的。

第三类“知道怎样做的知识 (Know - how)”, 指做某些事情的技艺和能力, 被称为技术情报和商业秘密, 其典型是企业开发和保存于其内部的技术诀窍或专有技术。

第四类“谁以及是怎样创造知识的 (Know - who)”, 侧重创造思想、方法、手段、过程以及特点等的了解。

2. 知识的来源

第一类是直接源于产生信息的客观事物。

第二类是通过信息载体或媒介 (文献、电视、广播等) 的传递、交流而间接获得的。

三、文献

文献是记录知识的一切载体。

目前有关文献的较权威的定义主要有两个: 一是《文献情报术语国际标准 (草案)》(ISO/DIS127) 的定义, 即“为了把人类知识传播开来和继承下去, 人们用文字、图形、符号、声频、视频等手段将其记录下来, 或写在纸上, 或晒在蓝图上, 或摄制在感光片上, 或录到唱片上, 或存贮在磁盘上。这种附着在各种载体上的记录统称为文献。”另一个是中国的国家标准, 是各种媒介和形式的信息集合, 包括文字、声像印刷品、电子信息、数据库等。

1. 文献的构成要素

(1) 文献信息是文献的内容

(2) 符号系统是信息的携带者

文字已不再是表达思想的唯一手段，图形、音频、视频等同样成为表达思想、传递感情的重要手段。总的说来，文献中的符号系统指图画、文字、公式、图表、编码、声像和电磁信息等。

(3) 载体是符号赖以依附的“寄主”

随着生产力的发展，存贮和表达人们思想的物质载体不同了，从金石、竹简、羊皮、丝帛、纸张发展到用感光介质和磁性介质。从非人工材质的泥版、石器、石板、兽骨、木板、竹片、兽皮、树叶、桦树皮等到各种人工材质，如无机材质（陶、砖瓦、瓷、玻璃等）、金属材质（青铜、铁、铝、金银等）、高分子材质（帛、纸、胶片、聚酯纤维等）、复合材质（磁带、光盘等）。纸质文献已经变得不便于快速传输信息、高效查阅和高密度存贮了，于是其他介质的文献应运而生，如光盘等。但纸质文献具有价格低廉、质地柔软、易于书写、携带和收藏等其他一些载体所无法比拟的性能，仍然是文献家族的主干。

(4) 记录方式是将文献的符号进入载体的方法和过程

按记录方法分手工记录、机械记录、光记录、电记录、声记录 and 磁记录。其中最常见的是印刷。随着科技的发展，文字记录可以转变为数据记录，并用电子方法存贮到磁介质上去。此外，还有光学字符识别（Optical Character Recognition，简称 OCR）技术就代表了这种崭新的技术，它通过光学方法对字符、标记表示的书面数据进行自动识别，转换成机器可以处理的信息，这样就实现了纸质文献信息向电子信息的转换，而且可以通过不同接口装置向不同设备输入数据，这些设备进而与通信网络相连，实现远距离数据传送。

2. 文献的类型

按文献载体可分：印刷型文献、缩微型文献、机读型文献、声像型文献

按文献加工程度可分：一次文献、二次文献、三次文献。

按文献的出版形式可分：有图书、期刊论文、会议文献、科技报告、专利文献、学位论文、标准文献、政府出版物、产品样本、科技档案等。

四、信息、知识与文献的关系

1. 区别

信息，有形的、独立于行动和决策、经过处理改变形态、物质产品、与环境无关、可以复制；知识，无形的、与行动和决策相关、经过处理改变思维、精神产品、环境改变涵义、经过学习才能转让、无法复制。

2. 联系

知识是有组织的大量的信息，获得知识有赖于获得信息；信息是知识得以形成和传播的中介，而不是知识本身，知识是经过精心研究、领会后的有用信息，是人类对信息加工处理后的产物。并非所有信息都可成为知识，在信息时代，源于众多复杂客体的大量信息，只有借助于现代化的信息手段，并通过掌握现代信息科学技术的认知主体，才能真正转化为知识。

第二节 信息检索

一、信息检索的涵义

信息检索的概念有狭义和广义之分。狭义的检索 (Retrieval) 是指依据一定的方法，从已经组织好的大量有关信息集合中，查找并获取特定信息的过程。这里的信息集合，不是通常所指的信息本身，而是关于信息的信息或线索。如果真正要获取信息所记录的内容，那么还要依据检索所取得的线索索取原文。

广义的检索包括信息的存储和检索两个过程 (Storage and Retrieval)。信息存储是指工作人员将大量无序的信息集中起来，根据信息源的外表特征和内容特征，经过整理、分类、浓缩、标引等处理，使其系统化、有序化，并按一定的技术要求建成一个具有检索功能的工具或检索系统，供人们检索和利用。而检索是

指运用编制好的检索工具或检索系统，查找出满足用户要求的特定信息。

二、信息检索的基本原理

信息检索的基本原理是：通过对大量的、分散无序的信息进行搜集、加工、组织、存储，建立各种各样的检索系统，并通过一定的方法和手段使存储与检索这两个过程所采用的特征标识达到一致，以便有效地获得和利用信息源。其中存储是为了检索，而检索又必须先进行存储。

文献信息的存储和检索的全过程可用图 1-1 表示。

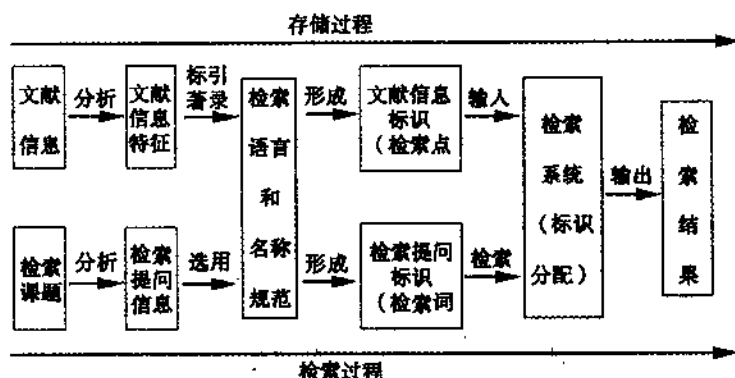


图 1-1 文献检索基本原理示意图

存储过程主要是对信息源进行标引，将其外表和内容特征（如文献的标题、作者、来源和主题等）用特定的检索语言转化为一定的标识（如主题词、分类号和类目名称等），再将这些标识按一定的顺序编排后输入检索系统，从而为检索提供有规可循的途径。

为了保证文献信息能存得进、取得出，就必须使文献存储所依据的规则与文献信息检索所依据的规则尽量做到一致。也就是说，为了检索过程的顺利进行和达到较高的检索效率，除了在存储和检索过程的各个环节必须依据一定的方法和规则外，还必须

有统一的检索语言和名称规范作为存储人员和检索人员的共同依据。

三、信息检索系统

信息检索是通过对信息系统的操作来实现的，信息系统包含信息集合的载体、技术设备和语义规则。按信息的存储媒体和技术手段来划分，把信息系统分为两大部分：手工检索系统和计算信息检索系统。

1. 手工检索系统 (Manual Retrieval System)

手工检索是计算机检索的基础和先声。现在计算机检索的基本理论和检索方法都是从手工检索发展而来的。而且，目前手工检索工具在所有的检索工具中仍占有相当大的比重，某些学科和科研课题的检索利用手工检索也较为方便。同时，手工检索与计算机检索相比查准率更高，其过程也不需要额外费用。所以，了解重要的手工检索工具及其检索方法仍必要。

手工检索工具的组成主要由三部分组成：

(1) 文摘部分：这部分是对收录文献的详细介绍，包括文摘号、文献标题、作者、作者工作单位、文献来源（如期刊、会议录、专利等）、文摘。

(2) 索引部分：这部分将所有文献按一定的文献特征、（如主题词、分类号、作者等）以一定的次序（如字母顺序、拼音顺序等）排列，每一个文献特征对应一个或几个文摘号或其他号码。从索引中获取的文摘号等可以过度到文摘部分从而了解该文献。重要的索引包括主题索引、分类索引和作者索引等。

(3) 附录部分：包括主题词表（供选择主题词用）、缩略语表、期刊来源索引等。

有些检索工具将前两部分放在一起（如《工程索引》），有些索引的部分与文摘部分分开（如《化学文摘》的卷索引），有些索引对文献不做详细介绍（如《最新目次》中收录的文献就没有文摘），也有些索引没有附录。

2. 计算机检索系统 (Computer-based Retrieval)

与手工检索相比,计算机检索的特点是:

(1) 速度快、效率高,仅几分钟就可以从成千上万条记录中找出所需信息。

(2) 检索范围广,可以迅速而方便地浏览相关学科或主题的所有数据库中的记录,在网络中,几乎每一台个人计算机都可以成为信息源。

(3) 检索不受时空的限制,只要拥有相应的软件和硬件设备,就可以在任何地方借助光盘和通信网络查询所需信息。

(4) 由于数据更新快,可以及时获得最新信息。

(5) 检索辅助功能完善,使用方便,检索软件可采用菜单驱动,几乎所有检索系统都有查询服务或提供操作演示盘,界面友好的程度在不断提高,检索结果的输出方式丰富多样,并可以按要求做排序、统计、绘图等加工。

在后面的章节对计算机检索有深入论述,这里不再多谈。

四、信息检索的方法和策略

1. 检索方法

检索方法是为实现检索计划或方案所提出的检索目的而采取的具体操作方法或手段的总称。检索获取知识信息的方法主要有两种,即直接检索和间接检索。

直接检索是通过浏览各种出版物上发表的学术论文、专利说明书等,以了解有关学科或专题发展动态的一种最简单的检索方法。其优点是能立即明确判断所包含的知识信息是否具有针对性和实用价值,不足是存在较大的盲目性和偶然性,查全率较低。

间接检索是借助于各类检索系统,从数量庞大的信息集合中,迅速、准确地查找特定课题的有关知识信息的常用检索方法,其优点是所获得知识信息的全面性和准确性都较高,这也是人们普遍采用的方法。

间接检索方法的全过程,一般可分为信息需求分析、制定检索策略、实施检索策略和评价检索效果。在此,主要对检索策略做一个详细地解释。