

INFORMATION RETRIEVAL

国家精品课系列教材  
国家级精品资源共享课教材  
教育部大学计算机课程改革项目成果

# 网络信息检索与利用

Information Retrieval and  
Utilization on Network

董卫军 高 飞 编著  
耿国华 主审



电子工业出版社  
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY  
<http://www.phei.com.cn>

014035852

G254.92  
18

国家精品课教材  
国家级精品资源共享课教材  
教育部大学计算机课程改革项目成果

# 网络信息检索与利用

董卫军 高 飞 编著

耿国华 主审



电子工业出版社

G 254.92

Publishing House of Electronics Industry

北京·BEIJING

18



北航

C1723252

## 内 容 简 介

本书是教育部大学计算机课程改革项目成果之一，也是国家精品课和国家级精品视频共享课“计算机基础”的辅助教材。教材立足于“面向实践，突出应用”，满足广大信息用户快速、准确、有效地获取有效网络信息资源的需求。

本书理论与实践相结合，课题查新案例分析与科技创新引导交融，共7章，包括：信息检索概述，网络基础与网络信息安全，Internet 信息检索，常见中文数据库的使用，常见外文数据库的使用，PDF 文件与 CAJ 文件的处理，论文的撰写。在介绍基本理论的基础上，每章内容设计了大量的检索示例，便于学习者快速掌握相关技术。

本书可作为高等学校各专业“网络信息资源检索”课程的教材，也可以作为各类计算机应用技能社会培训课程教材和广大科研工作者掌握网络信息检索方法与技巧的实用参考工具书。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。  
版权所有，侵权必究。

### 图书在版编目(CIP)数据

网络信息检索与利用 / 董卫军, 高飞编著. —北京: 电子工业出版社, 2014.4  
ISBN 978-7-121-22629-8

I. ①网… II. ①董… ②高… III. ①网络检索—高等学校—教材 IV. ①G354.4

中国版本图书馆 CIP 数据核字 (2014) 第 045286 号

策划编辑: 任欢欢

责任编辑: 章海涛

特约编辑: 何 雄

印 刷: 三河市鑫金马印装有限公司

装 订: 三河市鑫金马印装有限公司

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本: 787×1092 1/16 印张: 12 字数: 340 千字

印 次: 2014 年 4 月第 1 次印刷

定 价: 32.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888。

质量投诉请发邮件至 [zltz@phei.com.cn](mailto:zltz@phei.com.cn), 盗版侵权举报请发邮件至 [dbqq@phei.com.cn](mailto:dbqq@phei.com.cn)。

服务热线: (010) 88258888。

# 前 言

信息资源与材料资源、能源资源共同构成经济和社会发展的三式战略资源。互联网的广泛应用彻底改变了人类社会获取、处理与利用信息的方式和效率。然而，在海量信息中存在着大量虚假无用信息，这使得获取有用的信息资源变得越来越困难。因此，人们不仅需要多种知识，更需要掌握发现知识、过滤知识、分析利用知识的能力。作为新时代的的大学生，信息检索能力的培养显得尤为重要，这不仅是现代人才的必备素质，也是一个人知识和能力可持续发展的重要条件。

本书是国家精品课程、国家精品视频共享课“计算机基础”的辅助教材，也是教育部大学计算机课程改革项目——“基于计算思维的大学计算机系列课程及教材建设”项目的成果之一。本书是在总结信息检索与利用课程的教学实践和经验，体现计算思维教学改革的基础上，以教育部高等学校大学计算机课程教育教学指导委员的基本要求为指导而精心撰写的。

本书立足于“面向实践，突出应用”，课题查新案例分析与科技创新引导交融，满足广大信息用户快速、准确、有效地获取网络信息资源的需求。

全书共7章，包括：信息检索概述，网络基础与网络信息安全，Internet 信息检索，常见中文数据库的使用，常见外文数据库的使用，PDF 文件与 CAJ 文件的处理，论文的撰写。在介绍基本理论的基础上，每章内容中设计了大量的检索示例，便于学习者快速掌握通过网络和文献数据库获取有效信息的方法和手段，为自主学习和兴趣研究提供支持。

教材突出技术性、应用性与示范性，优先注重内容在应用上的层次性，适当兼顾整体在理论上的系统性。本书可作为高等学校“网络信息资源检索”课程的教材，也可作为“大学计算机”课程的辅助教材，还可作为各类计算机应用技能社会培训课程教材和广大科研工作者掌握网络信息检索方法与技巧的实用参考工具书。

本书由董卫军编写，由西北大学耿国华教授主审。在成书之际，感谢教学团队成员的样例支持。由于水平有限，书中难免有不妥之处，恳请指正。

本书为读者提供相关教学资料，可从华信教育资源网站(<http://www.hxedu.com.cn>)下载。

作 者

# 目 录

<b>第 1 章 信息检索概述</b> .....	<b>1</b>
1.1 信息概述.....	1
1.1.1 信息.....	1
1.1.2 文献.....	4
1.2 信息检索.....	7
1.2.1 信息检索的概念.....	7
1.2.2 信息检索系统.....	7
1.2.3 计算机检索.....	11
1.2.4 检索效果的评价.....	14
1.3 信息检索的基本步骤.....	15
1.3.1 分析研究课题.....	15
1.3.2 制订检索策略.....	16
1.3.3 选择检索工具.....	17
1.3.4 确定检索途径与选定检索方法.....	18
1.3.5 查找文献.....	18
1.3.6 检索举例.....	19
习题 1.....	20
<b>第 2 章 网络基础与网络信息安全</b> .....	<b>22</b>
2.1 网络基础.....	22
2.1.1 计算机网络的产生和发展.....	22
2.1.2 计算机网络的基本概念.....	23
2.2 计算机网络的基本组成.....	24
2.2.1 联网计算机.....	24
2.2.2 通信线路.....	25
2.2.3 通信设备.....	28
2.2.4 网络软件.....	30
2.3 因特网基础.....	31
2.3.1 因特网体系结构.....	31
2.3.2 IP 地址.....	33
2.3.3 域名系统.....	35
2.4 因特网基本服务.....	36
2.4.1 WWW 服务.....	37
2.4.2 电子邮件服务.....	38
2.4.3 文件传输服务.....	39
2.4.4 远程登录服务.....	40

2.5	代理服务器	40
2.5.1	代理服务器简介	40
2.5.2	代理服务器的获取与设置	42
2.6	网络安全	43
2.6.1	网络安全的含义与特征	43
2.6.2	网络安全攻击	43
2.6.3	基本网络安全技术	44
2.6.4	计算机病毒防治	49
	习题 2	58
<b>第 3 章</b>	<b>Internet 信息检索</b>	<b>60</b>
3.1	Internet 信息资源	60
3.1.1	Internet 信息资源特点	60
3.1.2	Internet 信息资源分类	60
3.1.3	Internet 信息资源的组织	62
3.1.4	影响 Internet 信息检索的因素	66
3.2	搜索引擎	67
3.2.1	搜索引擎概述	67
3.2.2	基本搜索技巧	70
3.2.3	常见搜索引擎简介	71
3.3	Google 信息检索	73
3.3.1	Google 的基本特点	74
3.3.2	基本搜索	74
3.3.3	字段搜索	77
3.3.4	高级搜索与学术搜索	79
3.3.5	其他重要功能	80
3.4	百度信息检索	81
3.4.1	主要服务介绍	81
3.4.2	基本检索	82
3.4.3	特殊检索	82
3.4.4	百度特色搜索	83
3.5	查询策略	83
3.5.1	合理选择查询关键词	83
3.5.2	综合使用各种查询方法	86
3.6	因特网文件下载	87
3.6.1	文件下载原理	87
3.6.2	利用 IE 浏览器下载	88
3.6.3	利用 BitComet 下载文件	88
3.6.4	利用迅雷下载文件	90
	习题 3	94

<b>第 4 章 常见中文数据库的使用</b> .....	<b>96</b>
4.1 常见中文数据库 .....	96
4.2 中文科技期刊数据库的使用 .....	97
4.2.1 中文科技期刊数据库简介 .....	97
4.2.2 基本功能 .....	99
4.2.3 基础检索 .....	99
4.2.4 高级检索 .....	102
4.3 中国知网数据库的使用 .....	104
4.3.1 中文知网数据库简介 .....	104
4.3.2 初级检索 .....	106
4.3.3 高级检索 .....	107
4.4 万方数据资源系统的使用 .....	108
4.4.1 万方数据库简介 .....	108
4.4.2 初级检索 .....	110
4.4.3 高级检索 .....	110
4.5 超星数字图书馆的使用 .....	111
4.5.1 超星数字图书馆的特点 .....	111
4.5.2 超星数字图书馆全文检索 .....	112
习题 4 .....	115
<b>第 5 章 常见外文数据库的使用</b> .....	<b>118</b>
5.1 常见外文数据库简介 .....	118
5.1.1 常见的美国数据库 .....	118
5.1.2 其他国家常见数据库 .....	121
5.1.3 三大检索工具 .....	122
5.2 Web of Science .....	123
5.2.1 Web of Science 简介 .....	123
5.2.2 Web of Science 的使用 .....	124
5.2.3 Web of Science 应用举例 .....	128
5.3 ISTP 的使用 .....	132
5.3.1 ISTP 简介 .....	132
5.3.2 检索方法 .....	132
5.3.3 检索结果的处理 .....	134
5.4 EI 的使用 .....	134
5.4.1 EI 简介 .....	135
5.4.2 EI Compendex Web 检索方法 .....	135
习题 5 .....	144
<b>第 6 章 PDF 文件和 CAJ 文件的处理</b> .....	<b>147</b>
6.1 PDF 文件处理 .....	147

6.1.1	PDF 文件简介	147
6.1.2	Adobe Reader 的使用	148
6.1.3	Foxit PDF Editor 的使用	152
6.1.4	通过 Word 生成 PDF 文件	158
6.2	CAJ 文件处理	159
6.2.1	CAJViewer 阅读器基本功能	159
6.2.2	CAJViewer 的使用	159
	习题 6	162
<b>第 7 章</b>	<b>论文的撰写</b>	<b>163</b>
7.1	科技论文的常见形式	163
7.1.1	学术论文、科技报告和专题研究论文	163
7.1.2	学位论文	164
7.1.3	简报、综述和评论	165
7.1.4	设计计算、理论分析和理论推导	165
7.2	科技论文的撰写	166
7.2.1	科技论文的特点	166
7.2.2	科技类论文的基本结构	166
7.2.3	科技论文的基本写法	168
7.2.4	摘要的书写	171
7.2.5	如何面对审稿意见	174
7.3	学士学位论文的撰写	174
7.3.1	学士学位论文的特点	174
7.3.2	学士学位论文的选题	175
7.3.3	学士学位论文的开题	177
7.3.4	学士学位论文的书写	177
7.3.5	学位论文写作应注意的问题	181
7.3.6	学位论文答辩	182
	习题 7	183
	<b>参考文献</b>	<b>184</b>



# 第1章 信息检索概述

信息已成为人类最重要的资源之一,如何获取信息、掌握信息和利用信息已经成为人们学习、工作和生活的基本技能,是否善于获取信息、掌握信息也成为影响成败的关键因素之一。

## 1.1 信息概述

### 1.1.1 信息

#### 1. 信息的概念

信息广泛存在于自然界、生物界和人类社会之中。不同事物有不同的本质、特征和运动规律,事物的特征通过一定的媒介或传递方式(如信号、声波、电磁波、图像等)使其他事物感知。从本质上讲,信息是事物存在的方式和运动状态及其规律的表现,是事物的一种普遍属性。对人类来说,信息与空气、水、食物等一样,都是不可缺少的。

#### 2. 信息的特征

信息具有以下基本特征。

① 客观性。信息客观存在,不以人的意志为转移。信息是现实世界中各种事物运动与状态的反映,可以被人们感知、处理、存储和使用。信息的客观性还表现在反映客观世界变化的信息包含在各种物质之中。

② 时效性。客观事物总是不断地发展变化,因而信息也会发展变化,如果信息不能适时地反映事物存在的方式和运动状态,那么该信息就失去其效用。

③ 传递性。信息的传递性是指任何信息只有从信源出发,经过信息载体传递才能被信宿接收并进行处理和运用。也就是说,信息可以在时间上或空间上从一点移动到另一点,可以通过语言、动作、文字、通信、计算机等多种渠道和媒介传播。

④ 共享性。同一内容的信息可以在同一时间或不同时间里被多个信息用户使用。一条信息被用户吸收和利用并不影响信息的本身和被其他用户的反复使用,各用户分享的份额也不会因为分享人数而受影响。

#### 3. 信息的常见类型

按照信息资源的开发程度,信息可分为潜在信息资源与现实信息资源两大类。现实信息资源又可分为口语信息资源、肢体语信息资源、实物信息资源、文献信息资源、网络信息资源和多媒体信息资源。

① 口语信息资源。口语信息资源是以交谈、聊天、授课、讨论等方式传递的信息资源。其特点是传递迅速,互动性强,但稍纵即逝、久传易出差异。因此,通过这种方式了解到的信息应记录下来,并加以证实。

② 肢体语信息资源。肢体语信息资源是以手势、表情、姿势(如舞蹈、体育比赛、杂技)等方式传递的信息资源。其特点是直观性强、生动丰富、印象深刻、富有感染力,但此类信息的容量有限。

③ 实物信息资源。实物信息资源是以诸如文物、产品样本、模型、碑刻、雕塑等实物形式表示的信息资源。其特点是直观性强、感觉实在、信息量大，但需要通过知识、智慧、经验和工具挖掘大量隐含的信息。

④ 文献信息资源。文献信息资源是以文字、图形、符号、音频、视频等方式记录在各种载体上的知识和信息，包括图书、连续出版物、小册子以及学位论文、专利、标准、回忆录、政府出版物等特种文献。其特点是经过加工、整理，较为系统、准确、可靠，便于保存和利用，但存在信息相对滞后、部分信息尚待证实的情况。

#### 4. 信息加工与数据挖掘

信息加工是对收集来的信息进行去伪存真、去粗取精、由表及里、由此及彼的加工过程，是在原始信息的基础上，生产出价值含量高、方便用户利用的二次信息的活动过程。这一过程将使信息增值，只有对信息进行适当处理，才能产生新的、用以指导决策的有效信息或知识。

随着信息技术的高速发展，人们积累的数据量急剧增长，如何从海量的数据中提取有用的知识成为当务之急。数据挖掘就是为顺应这种需要应运而生发展起来的数据处理技术。其主要任务是关联分析、分类、预测时序模式和偏差分析等。

##### (1) 数据挖掘 (data mining) 的现实需求

随着数据库技术的迅速发展以及数据库管理系统的广泛应用，人们积累的数据越来越多。激增的数据背后隐藏着许多重要的信息，人们希望能够对其进行更高层次的分析，以便更好地利用这些数据。

##### (2) 支持数据挖掘技术的基础

数据挖掘技术是人们长期对数据库技术进行研究和开发的结果。起初，各种商业数据是存储在计算机的数据库中的，然后发展到可对数据库进行查询和访问，进而发展到对数据库的即时遍历。数据挖掘使数据库技术进入了一个更高级的阶段，不但能对过去的数据进行查询和遍历，而且能够找出过去数据之间的潜在联系，从而促进信息的传递。

##### (3) 数据挖掘的概念

数据挖掘是从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中，提取隐含在其中的、人们事先不知道的、但又是潜在有用的信息和知识的过程。与数据挖掘相近的同义词有数据融合、数据分析和决策支持等。

数据挖掘包括以下含义：① 数据源必须是真实、海量、含噪声的；② 发现的是用户感兴趣的知识；③ 发现的知识要可接受、可理解、可运用；④ 发现的知识仅支持特定的发现问题。

因此，数据挖掘是一门交叉学科，把人们对数据的应用从低层次的简单查询提升到从数据中挖掘知识，提供决策支持。

##### (4) 数据挖掘的任务

数据挖掘的任务主要是关联分析、聚类分析、分类、预测、时序模式和偏差分析等。通常所说的数据挖掘会通过这些任务中的一个或者多个来对数据进行处理。

① 预测任务。这些任务的目标是根据其他属性的值，预测特定属性的值。被预测的属性一般称为目标变量或因变量，而用来做预测的属性称为说明变量或自变量。

描述任务的目标是导出概括数据中潜在联系的模式（相关、趋势、聚类、轨迹和异常）。本质上，描述性数据挖掘任务通常是探查性的，并且常常需要后处理技术验证和解释结果。

预测建模涉及以说明变量函数的方式为目标变量建立模型。预测建模任务有两类：分类和回

归。分类用于预测离散的目标变量，回归用于预测连续的目标变量。例如，预测一个 Web 用户是否会在网上书店买书是分类任务，因为该目标变量是二值的，而预测某股票的未来价格则是回归任务，因为价格具有连续值属性。两项任务目标都是训练一个模型，使目标变量预测值与实际值之间的误差达到最小。预测建模可以用来确定顾客对产品促销活动的反应，预测地球生态系统的扰动，或根据检查结果判断病人是否患有某种疾病。

② 关联分析 (association analysis)。两个或两个以上变量的取值之间存在某种规律性，就称为关联。数据关联是数据库中存在的—类重要的、可被发现的知识。关联分为简单关联、时序关联和因果关联。关联分析用来发现描述数据中强关联特征的模式，所发现的模式通常用蕴含规则或特征子集的形式表示。由于搜索空间是指指数规模的，关联分析的目标是以有效的方式提取最有趣的模式。关联分析的应用包括找出具有相关功能的基因组、识别用户一起访问的 Web 页面、理解地球气候系统不同元素之间的联系等。

③ 聚类分析 (clustering analysis)。聚类是把数据按照相似性归纳成若干类别，同一类中的数据彼此相似，不同类中的数据相异。聚类分析可以建立宏观的概念，发现数据的分布模式以及可能的数据属性之间的相互关系。聚类分析旨在发现紧密相关的观测值组群，使得与属于不同簇的观测值相比，属于同一簇的观测值相互之间尽可能类似。聚类可用于对相关的顾客分组、找出显著影响地球气候的海洋区域以及压缩数据等。

④ 异常检测 (anomaly detection)。异常检测的任务是识别其特征显著不同于其他数据的观测值。这样的观测值称为异常点或离群点。异常检测算法的目标是发现真正的异常点，而避免错误地将正常的对象标注为异常点。换言之，一个好的异常检测器必须具有高检测率和低误报率。异常检测的应用包括检测欺诈、网络攻击、疾病的不寻常模式、生态系统扰动等。

#### (5) 数据挖掘的过程

从数据本身来考虑，通常数据挖掘需要有信息收集、数据集成、数据规约、数据清理、数据变换、数据挖掘实施过程、模式评估和知识表示等 8 个步骤，如图 1.1 所示。

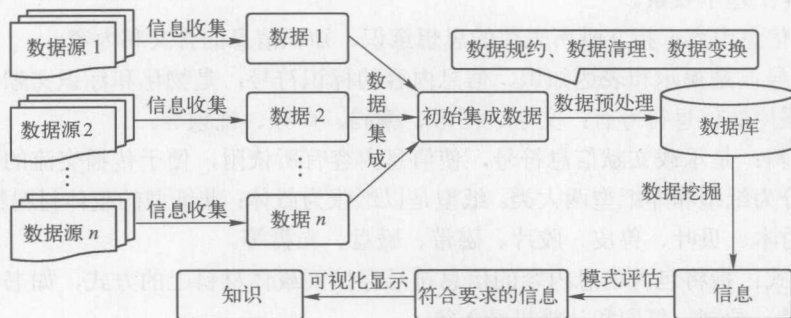


图 1.1 数据挖掘的基本过程

① 信息收集：根据确定的数据分析对象抽象出在数据分析中所需要的特征信息，然后选择合适的信息收集方法，将收集到的信息存入数据库。对于海量数据，选择一个合适数据仓库是至关重要的。

② 数据集成：把不同来源、格式、特点性质的数据在逻辑上或物理上有机地集中，从而为企业提供全面的数据共享。

③ 数据规约：数据挖掘算法即使在少量数据上也需要很长的时间，而做商业运营数据挖掘时往往数据量非常大。数据规约技术可以用来得到数据集的规约表示，规约表示数据量明显少于

原始数据，但仍然接近于保持原数据的完整性，并且规约后执行数据挖掘结果与规约前执行结果相同或几乎相同。

④ 数据清理：在数据库中的数据有一些是不完整的（有些感兴趣的属性缺少属性值）、含噪声的（包含错误的属性值），并且是不一致的（同样的信息不同的表示方式），因此需要进行数据清理，将完整、正确、一致的数据信息存入数据仓库，否则挖掘的结果会差强人意。

⑤ 数据变换：通过平滑聚集、数据概念化、规范化等方式，将数据转换成适用于数据挖掘的形式。对于有些实数型数据，通过概念分层和数据的离散化来转换数据也是重要的一步。

⑥ 数据挖掘：根据数据仓库中的数据信息，选择合适的分析工具，应用统计方法、事例推理、决策树、规则推理、模糊集甚至神经网络、遗传算法的方法处理信息，获得有用的分析信息。

⑦ 模式评估：从商业角度，由行业专家来验证数据挖掘结果的正确性。

⑧ 知识表示：将数据挖掘所得到的分析信息以可视化的方式呈现给用户，或作为新的知识存放在知识库中，供其他应用程序使用。

数据挖掘过程是一个反复循环的过程，如果没有达到预期目标，每个步骤都需要回到前面的步骤，重新调整并执行。当然，不是每个数据挖掘的工作都需要每个步骤，如在某个工作中不存在多个数据源的时候，数据集成的步骤便可以省略。

数据规约、数据清理、数据变换又合称数据预处理。在数据挖掘中，至少 60% 的费用可能要花在信息收集阶段，至少 60% 以上的精力和时间是花在数据预处理上。

## 1.1.2 文献

文献是记录信息和知识基本载体。狭义文献主要指具有历史意义的书面材料，广义文献是指记录知识的一切载体。

### 1. 文献的构成要素

文献包含 4 个基本要素。

① 知识、信息内容：指文献所表达的思想意识、知识信息的含义和内容。

② 信息符号：是揭示和表达知识、信息内容的标识符号，是物化和标识文献信息内容的工具。现代文献采用的信息符号有：文字、图形、编码、声频、视频等。

③ 载体材料：是承载文献信息符号，使信息内容有所依附，便于传播交流的物质材料。载体材料一般可分为纸型和非纸型两大类。纸型是以纸张为载体；非纸型的载体材料很多，如甲骨、金石、缣帛、竹木、贝叶、兽皮、胶片、磁带、磁盘、光盘等。

④ 记录方式：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、打字、印刷、拍摄、录制、复印和计算机录入等。

### 2. 文献的主要类型

文献的分类方法很多，按载体形式区分为印刷型（纸型）、机读型（电子型）、声像型（视听型资料）。按出版形式，文献可分为以下 7 类。

① 图书。图书是篇幅达到 48 页以上并构成一个书目单元的文献。其优点是内容全面而系统，观点较成熟，是求知解惑时主要利用的文献之一。其缺点是由于出版周期长，不能及时、迅速地反映最新科研成果。

② 期刊。期刊又称为杂志，是指定期或不定期的连续出版物。在期刊上可刊登多位作者的

文章, 设置若干栏目, 是求知或开展科学研究的重要文献信息源。电子期刊是利用计算机和通信系统进行撰写、编辑、传送的期刊形式。电子期刊具有报道速度快、出版周期短、数量大、品种多等特点。目前, 在科研人员利用的各类文献中, 期刊占 85% 以上。

③ 报纸。报纸按出版发行周期分, 报纸有日报、晚报、双日报、周报、旬报等。报纸的特点是传递迅速, 发行广泛, 信息量大, 政策性强。

④ 会议文献。会议文献是指在学术会议和专业会议上宣读或交流的论文、材料、讨论记录、会议纪要等文献, 能反映某些学科或领域的最新研究进展和成就, 具有较高的研究价值。会议文献分为会前文献和会后文献。会前文献包括会议征文通知、会议预印本、会议论文摘要等。会后文献出版形式多样, 有连续性会议文集、专题论文集、期刊论文、会议报告等。

⑤ 政府出版物。政府出版物是指各国政府部门及其专设机构出版的会议文件、法规、法令、政策、统计、调查报告等。政府出版物可分为行政性文献和科技性文献, 具有正式性和权威性, 对了解各国政治、经济、科技发展有着独特的参考作用。

⑥ 学位论文。学位论文是指高等院校、科研机构的毕业生和研究生, 为获取相应学位所提交的学术论文(硕士、博士论文等)。学位论文一般具有一定的独创性, 内容系统详尽, 是启迪思路、开创新研究领域的重要研究资料。

⑦ 档案。档案是社会活动、生产建设和技术工作中所形成文件的总称, 有“一手材料”、“历史凭证”之称。

### 3. 文献加工

文献加工是将大量分散、零乱、无序的文献进行整理、浓缩、提炼, 并按照一定的逻辑顺序和科学体系加以编排存储, 使之系统化, 以便于检索利用。

文献加工是情报部门的基础工作之一, 是建立文献数据库必不可少的前处理工作。文献加工包括: 著录、标引、编目以及题录、简介、文摘的编制。

#### (1) 著录

著录是对文献的各种特征进行分析、选择和记录的过程。通过著录, 将文献的各种特征反映在检索工具的各种载体上, 情报用户利用检索工具即可方便地了解和掌握所需要的特定文献。文献著录必须遵循一定的规则。

在国际上, 著录规则有《国际标准书目著录》(ISBD) 等。中国的著录规则有《文献著录总则》(GB3792.1—1983) 和《检索期刊条目著录规则》(GB3793—1983) 等。著录项目一般包括题名、责任者项、版本项、文献特殊细节项、出版发行项、载体形态项、丛编项、附注项、文献标准编号及有关记载项、提要项。

著录的载体形式有卡片、书本和数据库等。卡片式与书本式是传统手工方式著录。计算机应用后, 著录项直接输入计算机, 并利用计算机一次输入多次使用的功能, 存入数据机库, 同时打印编排卡片式、书本式检索工具。

#### (2) 标引

标引是对文献的内容特征进行分析, 并以标识符号(作为检索标识), 揭示该文献的工作, 最常用的为分类标引和主题标引。

分类标引是指以一定的观点为指导, 按文献学科知识内容, 参考用户需求以及文献形式体裁, 依据一定的分类法, 选标分类号, 借以组织分类目录和索引。

主题标引是指以一定的观点为指导, 按文献所论述的主题内容, 参考用户需求以及文献形式

体裁,依据一定的主题词表,选标主题词,借以组织主题目录和索引。

由于分类法和主题法具有不同功能,中国很多情报部门同时采用两种方法,以满足用户从不同角度检索文献的要求。

分类标引与主题标引的工作程序包括 3 步:① 对文献论述的内容进行主题分析,通过分析提炼主题概念;② 依据事先选定的分类法与主题词表,从中选择和组配恰当的标识符号(分类法用的是分类号,主题法用的是经过规范化处理的词语);③ 按一定顺序将有关文献的题名或代号(如文摘号)列在该标识符号后面。

### (3) 编目

对文献进行著录和标引,形成一系列描述和揭示文献外表特征和内容特征的条目,并将这些条目有序排列成文献目录一类检索工具的工作。目录的种类很多,功能和排序方法也有所不同。

- ◎ 供用户随时查用的称为读者目录,供情报部门内部使用的称为公务目录。
- ◎ 按不同文献类型组成的目录有图书目录、期刊目录、科技报告目录、会议文献目录、学位论文目录、专利文献目录、技术标准目录、声像资料目录等。
- ◎ 按目录反映的文献收藏范围分,有馆藏目录、联合目录等。
- ◎ 按目录反映的文献内容分,有综合目录、专科目录、专题目录等。
- ◎ 按目录的不同编排方法分,有分类目录、主题目录、作者目录、书名目录等。
- ◎ 按目录的载体形式分,有卡片式、书本式及磁带式目录等。

### (4) 题录、简介、文摘的编制

题录、简介、文摘的编制是对每篇文献进行的不同深度的加工工作。

- ① 题录:仅根据文献题名、作者、出处等外表特征进行著录、标引的条目。
- ② 简介:除题录部分外,再对文献内容进行一般性介绍。
- ③ 文摘:对文献内容进行实质性描述。

将一系列题录、简介、文摘条目有序编排,并定期或不定期出版,就分别成为题录式检索刊物、简介式检索刊物和文摘杂志。再对文献进行深一步加工,即对相关的一系列原始文献进行分析研究和综合评论,从而形成综述报告和述评报告等。

## 4. 一次文献、二次文献和三次文献

从对文献的生产加工层次来看,文献可以分成一次文献、二次文献和三次文献。

① 一次文献。专著、报纸、期刊、专利文献、标准文献、会议文献、样本等成品文献属于一次文献,即人们对自然和社会信息进行首次加工而成的文字记载。这是文献信息源的主要部分,数量极为庞大,由于在内容上是分散的、无系统的,因此不便于管理和传播。

② 二次文献。为了控制文献,便于人们查找,对一次文献进行再加工,通过整理、提炼和压缩,并按其外部特征(题名、作者、文献物理特征)和内容特征序化,形成另一类新的文献形式——目录、书目、索引、文摘——就是二次文献。二次文献不是一次文献本身的汇集,而是一次文献特征的汇集,从而使人们方便地找到一次文献,或了解一次文献的内容。

③ 三次文献。利用二次文献,选择有关的一次文献再加以分析、综合而编写出来的第三个层次的文献形式——专题报告、综述、进展以及手册、百科全书、年鉴等工具书——就是三次文献。三次文献具有系统性、综合性、知识性和概括性的特点,便于人们高效率地了解某一领域的状况、动态、发展趋势和有关情况。

## 1.2 信息检索

### 1.2.1 信息检索的概念

信息检索起源于图书馆的参考咨询和文摘索引工作。随着计算机问世, 计算机技术逐步应用于信息检索领域, 并与信息检索理论紧密结合起来。20 世纪 60 年代到 80 年代, 在信息处理技术、通信技术、计算机和数据库技术的推动下, 信息检索在教育、军事和商业等领域高速发展, 得到了广泛的应用。

广义的信息检索是指将信息按一定的方式组织和存储起来, 并根据信息用户的需要找出有关的信息过程。狭义的信息检索仅指从信息集合中找出所需要的信息的过程, 相当于人们通常所说的信息查寻。

#### (1) 信息检索的过程

信息检索的实质是将描述用户所需信息的提问特征, 与信息存储的检索标识进行比较, 从中找出与提问特征(对信息的需求分析后, 从中选出能代表信息需求的主题词、分类号或其他符号)一致或基本一致的信息。现在最普遍的信息检索是计算机信息检索。

信息检索应包括两个主要方面: 信息标引和存储过程, 信息的需求分析和检索过程。

① 信息标引和存储过程。标引是用检索语言和分类号、主题词表示信息, 通过对大量无序信息资源进行标引处理, 使之有序化, 并按科学的方法存储组成检索工具或检索文档的过程。

② 信息的需求分析和检索过程。是指分析用户的信息需求, 利用组织好的检索系统, 按照系统提供的检索方法和途径检索有关信息的应用过程。

#### (2) 信息检索的意义

掌握各种文献的检索方法, 也就掌握了获取知识的窍门, 提高了综合能力, 对于充分发挥创造力将产生极大的帮助。

① 有助于知识更新。科技发展使得人类社会的产业结构正处在急剧变化之中, 大批知识密集型工业相继涌现, 边缘交叉学科大量出现, 知识老化现象也不断加重, 只有不断自学才能拓宽知识面, 适应信息时代的要求才能在未来的竞争中取得更大的主动权。

② 有助于发展教育。新技术革命也可称为“知识革命”, 由此带来的信息时代也应称为“知识时代”, 各种现代化技术将会渗透到所有行业和部门, 这就对当代教育提出了更高的要求, 必须加快教育信息化的步伐, 推进信息化教育。

③ 有助于科学研究。通过对有关文献进行全面的调查研究, 摸清国内外最新研究动态(什么人在做, 取得了什么成果, 尚存在什么问题), 从而避免重复劳动, 有所创新、有所前进, 提高科研的速度和效益。

### 1.2.2 信息检索系统

信息检索系统是指根据特定的信息需求而建立起来的一种有关信息搜集、加工、存储和检索的多功能开放系统, 其主要目的是为人们提供信息服务。任何具有信息存储与检索功能的系统, 均可以称为信息检索系统。

#### 1. 信息检索系统的分类

信息检索系统的划分有多种方法。

### (1) 按照检索的功能划分

按照检索的功能划分,信息检索系统可以分为书目检索系统和事实数据检索系统。

书目检索系统主要对某一研究课题的相关文献进行检索,其结果是获得一批相关文献的线索,检索作业的对象是检索工具。书目检索按检索手段可以分为手工书目检索和计算机书目检索。

事实数据检索系统用于各种事实或数据的检索,如查找某一词的解释,某人、某事件、某地名、某企业及其产品情况等,其结果是获得直接的、可供参考的答案。进行事实数据检索时,人们要使用各种参考工具,如词典、百科全书、年鉴、手册、名录、表谱或者使用相应的数据库。

### (2) 按照检索的手段划分

按照检索的手段划分,信息检索系统可以分为手工检索系统和计算机检索系统。

手工检索系统是以手工方式存储和检索信息的系统,检索时使用各种纸质工具,检索入口少、速度慢、效率较低。计算机检索系统是用计算机进行信息存储和检索的系统,检索时使用各种数据库,检索灵活、检索入口多、速度快、效率高。

## 2. 书目检索系统

按检索手段,书目检索可分为手工书目检索和计算机书目检索。

### (1) 检索工具的结构

检索工具主要由以下几部分组成。

① 说明:介绍检索工具的编辑方针、收录范围、总体结构、各部分的功能、体例及使用方法等,是使用检索工具的指南。

② 正文(主文档):指文献条目(也称为记录)的有序集合,是检索工具的主体。条目(记录)通常由文献的顺序号、篇名、著者、来源等著录项目(也称为字段)组成。为便于计算机识别,每个著录项目都设有字段标识符,常见的字段标识符如表 1.1 所示。

表 1.1 常见的字段标识符

字段名	标识符	全称	字段名	标识符	全称
文献题名	TI	Title	语种	LA	Language
摘要	AB	Abstract	国际标准刊号	IS	ISSN
叙词	DE	Descriptor	分类号	AF	Author affiliation
关键词	ID	Identifier	记录存储号	CP	Country of publication
文献类型	DT	Document type	出版年	PY	Publication year
著者	AU	Author	分类号	CC	Classification code
出版者	PB	Publisher	—	—	—

③ 索引。索引由各种辅助索引组成,在数据库中被称为索引文档。索引提供多种检索途径(检索入口),满足用户不同的检索需要,提高检索的速度和准确性。常见的辅助索引有著者索引、主题索引、文献号码索引等。

④ 附录。附录由来源期刊一览表、缩略语与全称对照表、不同语种音译对照表等组成。附录是正文的必要补充,有利于用户进一步理解正文的著录内容,了解文献来源情况,根据缩写查找全称等。

⑤ 词表。词表是检索系统中文献信息整序的依据,如分类表、主题词表等。词表通常独立存在,是准确选择检索分类号和主题词的依据。



## (2) 检索工具的类型

按照描述文献信息特征的方式不同、用途不同,检索工具可以分为目录、题录、文摘、索引。

① 目录。目录是以完整的出版单元(如一种图书、一种期刊)为单位,按照一定次序编排的对文献信息进行描述和报道的工具,也称为书目。目录对文献的描述比较简单,每条记录的字段主要包括文献题名、责任者、出版事项、分类号、主题词等。一种出版物经过如此描述后形成一条记录,将所有记录组织起来就形成了目录。

按照收录文献的类型,检索工具可分为:图书目录、期刊目录、标准目录、报纸目录、地图目录、档案目录等;按照收录文献的内容范围,可分为:综合目录、专题和专科目录、个人著述目录(亦称个人著述考,揭示与报道特定人物的全部著作以及有关该人物的文献)、地方文献目录(揭示与报道某地区文献的目录)等;按照功用划分,可分为:登记书目、书业书目、收藏目录、专题目录、推荐书目、书目之书目。

② 题录。题录是描述文献各种特征的一条记录,通常由文献的分类号、文献篇名、著者及其单位、文献来源、主题词等项组成。题录型检索工具(简称题录)是以单篇文献(如一篇论文、书中的一章)为报道单位。这是它与目录的主要区别,题录揭示文献比目录更详细。

③ 文摘。文摘是指通过描述文献的各种特征,并介绍文献主要内容来报道文献的工具。与题录型检索工具的报道单位一样,文摘也是单篇文献,只是著录项目(字段)比题录多了一个摘要项。文摘型检索工具(简称文摘)是书目检索系统的核心,是书目检索的主要工具。

④ 索引。索引是将某种信息集合(如目录、题录、文摘)中的一组相关信息,按照某种顺序组织并指引给用户的一种指南。索引具有便于检索,揭示文献信息深入、详细的优点。通常,索引从属于特定的信息集合,为其提供多种不同的检索途径(入口),在检索系统中占有重要地位。索引条目一般由检索标志和存储地址组成。标识是索引条目所指示的文献信息某方面的特征,如主题词、著者姓名、分类号等。存储地址是指明标识所表达的特定信息在信息集合中的地址,它通常是文献存取号(文摘号)。不同的标识系统构成不同的索引,常见的索引有以下几种。

- ◎ 分类索引:以分类号或类目名称作为索引标识,按照分类号排列形成的索引。
- ◎ 主题索引:以主题词(叙词或关键词)作为索引标识,按其字顺排列形成的索引。
- ◎ 著者索引:以文献上署名的著者、译者、编者等责任者的姓名或机关团体名称作为索引标识,按其字顺排列形成的索引。
- ◎ 专用索引:以某些领域专用的名词术语或符号作为索引标识编排形成的索引。
- ◎ 引文索引:以引文著者和引文的其余题录部分作为标识编制成的索引。

## 3. 事实数据检索系统

事实数据检索系统用于各种事实或数据的检索,如查找某一词的解释,某人、某事件、某地名、某企业及其产品情况等,其结果是获得直接的、可供参考的答案。进行事实数据检索时,使用各种参考工具,如词典、百科全书、年鉴、手册、名录、表谱、图录或者相应的数据库。表 1.2 描述了不同类型检索系统的用途。

① 字典/词典。字典/词典是以字/词为标目,着重解释字/词的读音、形体、意义及其用法,并按一定次序(通常为字顺)编排形成的工具。字典/词典主要回答有关词的读音、意义、用法、来源、演变,以及同义词、反义词、缩写等方面的问题。词典一般可划分为语文词典和专科词典两大类。通过网络,可以使用各种在线词典。