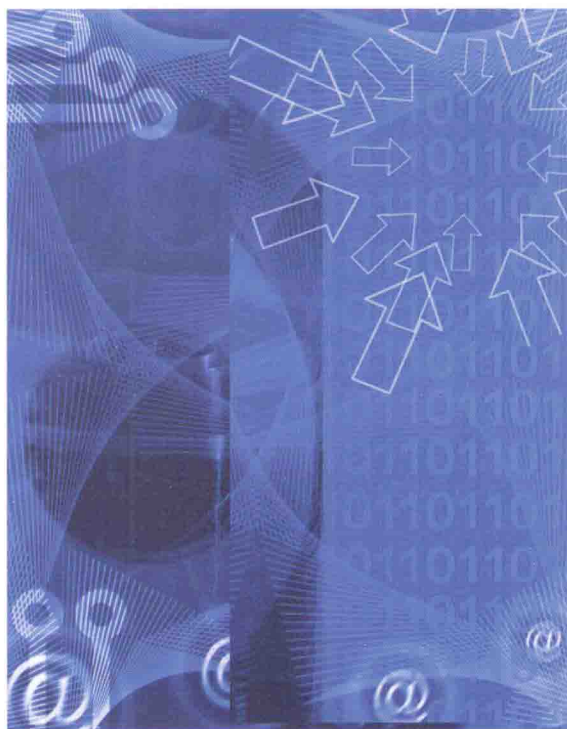


信息检索技术

- ◆ 信息检索的原理、方法
- ◆ 中国知网(CNKI)
- ◆ 中文数据库检索系统
- ◆ 搜索引擎应用
- ◆ 专利与标准文献检索
- ◆ 国外科技信息资源检索
- ◆ 信息资源的利用
- ◆ 文献信息服务



汪楠 成鹰 主编
孙颖 李淼焱 杨凯宁 副主编
张炎 主审



清华大学出版社

014058727

G254.97-43
10

高等学校计算机应用规划教材

信息检索技术

汪楠 成鹰 主编

孙颖 李森焱 杨凯宁 副主编



清华大学出版社



北航

C1745503

G254.97-43

10

内 容 简 介

本书是通用性很强的有关信息检索原理、方法和技能的著作,具有实用性、新颖性的特点。本书从信息检索的相关概念、基本原理以及常用技术和方法出发,注重常用检索工具的结构、检索方法和使用技巧的介绍,同时,借助大量的检索范例,使人们能有效地掌握现代网络环境下数字化信息资源的检索和利用方法。本书结构合理、内容全面、重点突出、注重实用性,既可作为高等院校信息检索课程的教学用书,又可作为教学、科研、工程技术和社会各界信息用户检索信息的参考书。

本书的电子教案和习题答案可以到 <http://www.tupwk.com.cn/download> 网站下载。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

信息检索技术 / 汪楠, 成鹰 主编. —北京: 清华大学出版社, 2014

(高等学校计算机应用规划教材)

ISBN 978-7-302-37465-7

I. ①信… II. ①汪… ②成… III. ①情报检索—高等学校—教材 IV. ①G252.7

中国版本图书馆 CIP 数据核字(2014)第 176170 号

责任编辑: 胡辰浩 马玉萍

装帧设计: 牛静敏

责任校对: 邱晓玉

责任印制: 沈 露

出版发行: 清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址: 北京清华大学学研大厦 A 座 邮 编: 100084

社 总 机: 010-62770175

邮 购: 010-62786544

投稿与读者服务: 010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈: 010-62772015, zhiliang@tup.tsinghua.edu.cn

课 件 下 载: <http://www.tup.com.cn>, 010-62794504

印 刷 者: 北京富博印刷有限公司

装 订 者: 北京市密云县京文制本装订厂

经 销: 全国新华书店

开 本: 185mm×260mm

印 张: 15.75

字 数: 364 千字

版 次: 2014 年 8 月第 1 版

印 次: 2014 年 8 月第 1 次印刷

印 数: 1~3000

定 价: 32.00 元

产品编号: 059990-01

前言

当今社会既是一个信息时代，又是一个网络时代。科学技术突飞猛进、社会竞争日益激烈、知识的无限性与个人能力的有限性的社会现实，已经摆在现代人的面前。如何在纷繁复杂的信息海洋中查找到适合的信息？掌握现代信息检索理论和信息检索技术的全新应用是人们的首选。

信息资源是人类进行学习、生活、科学研究和事业发展的基础，而信息资源的有效存储和检索作为人们获取信息资源的主要手段，也已经成为社会所有成员必须掌握的基本知识和技能。目前，“掌握资料查询、文献检索及运用现代信息技术获取相关信息的基本方法”已被规定为大多数专业的人才培养要求。可见加强信息检索课程的教学，增强学生的信息意识，重视培养学生的收集处理信息的能力、获取知识的能力，已经成为当今人才培养的重要任务之一。

当今社会的网络信息媒体层出不穷，信息检索技术不断更新，新兴的信息检索理念和方法不断涌现，为了适应社会的变化和教学的需要，本书的作者在编著了《信息检索方法与实践》、《数字信息检索》和《实用检索技术》之后，根据读者的反馈，推出了又一新作《信息检索技术》。

本书本着充实新内容、体现新进展、完善新体系的原则和目标，以信息资源的收集、分析、利用和再生为主线，将信息检索的知识体系进行了重组，使读者在掌握信息检索技术的同时，能更清晰地理解信息资源利用的意义，为终身学习奠定基础。

本书具有系统性、实用性和新颖性，从检索的基本原理出发，针对不同的信息资源的检索需求，注重检索方法和常用检索工具的介绍。本书为读者提供了大量的检索实例，详细介绍了检索分析、检索策略的制定、检索结果的甄别以及检索系统评价等，同时，进一步为读者提供了信息资源的收集、整理方法以及科技文献的写作方法。

本书共分7章，分别介绍了绪论、中国知网(CNKI)、搜索引擎应用、专利与标准文献检索、国外科技信息资源检索、信息资源的利用、文献信息服务等内容。本书融入了新的理念，注重实用性，理论与案例分析相得益彰，突出了检索实践环节，为读者提供了丰富的实训项目。本书设置了“本章小结”、“关键术语”、“知识窗”、“小提示”、“检索实训”等栏目，方便了学生学习信息检索知识和掌握信息检索技能。

本书由汪楠、成鹰担任主编，孙颖、李淼焱、杨凯宁担任副主编，张炎担任主审。王妍、于洁、方新儒、曹辉、周丹铭、陈秀明等参与了本书的校对和审读工作。

本书在编写过程中借鉴了许多网上资源和相关学者的研究成果，在此谨向相关机构和作者致以诚挚谢意。

由于作者的水平有限，书中疏漏和不足之处，恳请专家和读者批评指正。我们的电话是 010-62796045，信箱是 huchenhao@263.net。

作者

2014 年 5 月

目

第1章 绪论	1
1.1 信息与文献基本知识	1
1.1.1 从信息到文献	1
1.1.2 信息的分类	4
1.1.3 文献的分类	4
1.2 信息检索基础	8
1.2.1 信息检索的原理	8
1.2.2 信息检索的类型	9
1.2.3 信息检索的工具	10
1.2.4 信息检索的意义	14
1.3 信息检索的途径	15
1.3.1 外部特征途径	15
1.3.2 内容特征途径	16
1.4 信息检索的步骤和方法	17
1.4.1 信息检索的步骤	17
1.4.2 信息检索的方法	20
1.5 计算机信息检索基础	21
1.5.1 计算机信息检索概述	22
1.5.2 计算机信息检索技术	23
1.5.3 光盘检索	24
1.5.4 网络检索	25
本章小结	26
第2章 中国知网(CNKI)	30
2.1 中国知网(CNKI)综述	30
2.1.1 关于中国知网	30
2.1.2 中国知网数据资源	31
2.1.3 中国知网检索综述	32
2.2 中国知网初级检索	33
2.2.1 CNKI 初级单库检索	33
2.2.2 CNKI 初级跨库检索	35

录

2.3 中国知网高级检索	37
2.3.1 CNKI 单库高级检索概述	37
2.3.2 CNKI 跨库高级检索	41
2.4 中国知网其他检索	43
2.4.1 中国知网专业检索	43
2.4.2 中国知网出版物检索	46
2.5 常用中文数据库检索系统	49
2.5.1 万方数据	49
2.5.2 维普资讯网	52
2.6 检索实例	56
本章小结	60
第3章 搜索引擎应用	65
3.1 网络信息资源概述	65
3.1.1 网络信息资源检索的特点	66
3.1.2 搜索引擎概念及工作原理	67
3.1.3 搜索引擎检索的分类	69
3.1.4 搜索引擎的使用技巧	71
3.2 搜索引擎的利用	73
3.2.1 Baidu 及基本操作	74
3.2.2 地图搜索	77
3.2.3 图像搜索	83
3.3 移动搜索	88
3.3.1 移动搜索技术	88
3.3.2 移动搜索特点	90
3.3.3 移动搜索服务	91
3.3.4 移动搜索示例	92
3.4 搜索引擎选择	94
3.4.1 搜索引擎常见问题	94
3.4.2 搜索引擎的选择	95
3.5 检索实例	95
本章小结	99

第4章 专利与标准文献检索 105**4.1 专利文献 105**

4.1.1 专利基础知识 105

4.1.2 专利文献的情报价值 108

4.1.3 专利文献的分类方法及
检索途径 109**4.2 中文专利检索 115**4.2.1 国家知识产权局专利网络
检索系统 116

4.2.2 中国专利信息网 120

4.3 标准文献 121

4.3.1 标准文献基础知识 121

4.3.2 标准文献的商业价值 122

4.3.3 标准文献的分类方法 122

4.4 标准文献检索 125

4.4.1 国内标准文献检索工具 125

4.4.2 中国标准服务网检索 128

4.4.3 国际标准文献检索工具 130

4.4.4 ISO 国际标准的网络信息
检索 131**4.5 检索实例 133****本章小结 135****第5章 国外科技信息资源检索 140****5.1 美国《工程索引》 140**

5.1.1 《工程索引》概况 140

5.1.2 《工程索引》印刷版 141

5.1.3 《工程索引》光盘数据库 143

5.1.4 《工程索引》网络版 143

5.1.5 EI Compendex Web 检索
方法 144**5.2 美国《科学引文索引》 147**5.2.1 美国费城科学信息
研究所(ISI) 147

5.2.2 《科学引文索引》概述 147

5.2.3 《科学引文索引》印刷版 148

5.2.4 《科学引文索引》网络版 148

5.2.5 《科学引文索引》检索

方法 149

5.3 英国《科学文摘》 155

5.3.1 概述 155

5.3.2 《科学文摘》印刷版 156

5.3.3 《科学文摘》Web 版 157

5.3.4 《科学文摘》检索方法 158

5.4 美国《化学文摘》 159

5.4.1 《化学文摘》概况 159

5.4.2 CA 网络版——SciFinder 160

5.4.3 SciFinder 检索方法 161

5.5 检索实例 164**本章小结 174****第6章 信息资源的利用 179****6.1 信息资源的收集及分析 179**

6.1.1 信息资源的收集 179

6.1.2 信息资源的鉴别和整理 181

6.1.3 信息资源的分析 182

6.2 信息资源的利用 184

6.2.1 信息资源的再生 184

6.2.2 学术论文 184

6.2.3 文献综述 188

6.2.4 信息资源研究报告 190

6.3 信息资源研究的评价 192

6.3.1 信息资源的评价 192

6.3.2 信息资源利用效果评价 194

6.3.3 提高信息资源利用的方法 197

6.4 检索实例 200**本章小结 204****第7章 文献信息服务 208****7.1 文献信息服务系统 208**

7.1.1 图书馆系统 208

7.1.2 科技信息研究系统 212

7.1.3 档案馆系统 212

7.1.4 中国新闻出版信息网 213

7.2 图书资源检索	215	7.4.1 科技查新知识	230
7.2.1 常见的图书类型	215	7.4.2 科技查新的意义和作用	231
7.2.2 图书馆信息资源导航	216	7.4.3 科技查新流程	232
7.2.3 图书馆信息资源服务	217	7.4.4 科技查新规范	234
7.2.4 图书的检索途径	218	7.5 检索实例	234
7.3 中文全文数字图书馆	219	本章小结	235
7.3.1 中文全文数字图书馆概述	219	附录: 模拟试卷	240
7.3.2 超星数字图书馆	220	参考文献	243
7.3.3 其他数字图书馆	224		
7.4 科技查新	229		

第1章 绪 论

学习目标

1. 了解信息的特征、信息检索的概念、信息检索的目的和意义。
2. 理解检索工具的结构及目录、索引、文摘 3 种主要检索工具的特点及作用。
3. 熟悉信息检索的原理、类型和基本工具的使用。
4. 了解信息检索的基本途径，掌握信息检索的方法和步骤。

随着信息时代的到来以及全球市场的形成，信息的瞬息万变给人类社会带来了前所未有的机遇与挑战。在这样多变的环境下，利用具有潜在价值的数据资产，及时、准确地获取和传播有价值的信息，是人类社会存在和发展的前提条件之一。

1.1 信息与文献基本知识

未来学家托夫勒把人类社会的发展过程划分为 3 个阶段：以农业经济为基础的农业社会；以工业经济为基础的工业社会；以信息经济为基础的信息社会。在信息社会中，信息资源已被人们公认为是物质、能量资源之后的人类可利用的第三类资源，是人类社会领域的基础性资源。

1.1.1 从信息到文献

在当今的信息社会中，只有将自然现象和社会现象的信息上升为对自然和社会发​​展规律的认识，这种再生信息才能在社会和人类生活中发挥日益重要的作用。

1. 信息(Information)

美国数学家、控制论的主要奠基人维纳在 1950 年出版的《控制论与社会》中是这样来理解信息的：“人通过感觉器官感知周围世界”，“我们支配环境的命令就是给环境的一种信息”，因此，“信息就是我们在适应外部世界，并把这种适应反作用于外部世界的过程中，同外部世界进行交换的内容的名称”。“接收信息和使用信息的过程，就是我们适应外界环境的偶然性的过程，也是我们在这个环境中有效地生活的过程”。

从哲学的观点看，“信息”是物质的一种普遍的属性，它反映不同物质所具有的不同本质、特征以及运动状态和运动规律。它是用来消除人们对客观物质不确定性认识的东西。

“信息”是人们认识事物和获取新知识的唯一方式。只有掌握了事物的信息，人们才能对事物的运动规律进行调控，人们在认识世界和改造世界的过程中才能处于主动地位。

中国国家标准《情报与文献工作词汇——基本术语》(GB/T4894—1985)中信息的定义为：信息是物质存在的一种方式、形态或运动状态，也是事物的一种普遍属性，一般指数据、消息中所包含的意义，可以使消息中所描述事件的不确定性减少。

信息具备以下基本属性。

(1) 客观性。信息的客观性是指信息是事物本质特征和客观规律的表征，是客观存在的，不以人们的主观意识为转移。

(2) 传递性。信息在事物之间的相互联系必定在信息的流动中发生。信息的传递性表现在人与人之间的信息交换，人与计算机之间的信息交换，动物界及植物界之间的信号交换等方面，同时，人类进化过程中的细胞选择、遗传也是信息的传递与交换。

(3) 多态性。信息在不同的领域具有不同的特性或表现形式，如：客观事物中的各种自然属性；人工设备的技术特征；人类社会的各种社会特征；人脑中反映客观事物认识的思想、知识；人类交流信息过程中的声音、文字、图像以及用各种编码形式记录下来的数据、新闻、情报、消息等。各种形式的信息又常常以综合的方式表现事物的特征，所谓“多媒体”正是信息多样性和综合性的集中体现。

(4) 共享性。同一个信息资源可以被不同用户在不同地点、不同时间中同时利用，不需要任何限制条件，信息量也不会损失。信息不会像物质一样因为共享而减少，反而可以因为共享而衍生出更多信息。

(5) 时效性。信息的时效性是指信息的效能依赖于时间，它既表明了信息的时间价值，又表明了信息的经济价值和社会效益。客观事物不断变化而产生的信息具有很强的时效性。客观事物变化越快，信息的时效性就越强。时间的延误，就是扼杀信息的生命——导致信息价值衰减，乃至完全消失。时过境迁的信息是毫无价值的。

(6) 价值相对性。由于人们认识的能力与认知的条件不同，信息的接受者获得的信息及信息量也就不同，从这个意义上讲，信息的价值是有相对性的。

(7) 依附性。信息必须通过一定的符号，依附于某一载体才能被表现。没有载体就没有信息。

(8) 增长性。物资和能源在使用后会有消耗，而信息的使用不但不减少信息的数量，还能产生更多的信息。

2. 知识(Knowledge)

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，是对客观世界物质形态和运动规律的认识。根据韦伯斯特(Webster)词典 1997 年的定义，知识是通过实践、研究、联系或调查获得的关于事物的事实和状态的认识，是对科学、艺术或技术的理解，是人类获得的关于真理和原理的认识的总和。

从信息的观念看，知识来源于信息，是信息的一部分。简而言之，知识是理性化、优化、系统化了的的信息。从反映的内容而言，知识是客观事物的属性与联系的反映，是客观

世界在人脑中的相对正确的反映。从反映的活动形式而言,知识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。

知识是人们在实践活动中获得的关于世界的最本质的认识,是对信息的提炼、比较、挖掘、分析、概括、判断和推论。一般而言,知识具有共享性、传递性、非损耗性(可以反复使用,其价值不会减小)及再生性等特点。

3. 情报(Intelligence)

情报是作为交流对象的有用的知识。“情报是判断、意志、决心、行动需要的能指引方向的知识 and 智慧”;“情报是解决问题所需要的知识”;“情报是激活了的知识”。情报的基本属性是知识性、传递性和效用性。

(1) 知识性。一般而言,情报是一种新的知识。然而称为情报的知识并不排斥时间上的“不新”。这就是说,凡原先不知道的(不论时间多久),而现在又迫切需要的知识,都属于“新”的知识。

(2) 传递性。情报传递的过程包括从情报源进行收集到向用户提供情报的全过程。无论多高深、渊博、新奇的知识,如果不记录、不传递、不交流,都不能称为情报。

(3) 效用性(判断的标准)。情报是一种有价值、有效用的知识。情报均须“有用”,当用户不需要时,百分之百的知识也不能把它称之为情报。

4. 文献(Literature/Document)

国家标准《文献著录总则》(GB/T3792.1—1983)给出的定义为:“文献是记录有知识的一切载体。”国家标准化组织 ISO 在《文献情报术语国际标准》(ISO/DIS5217)给出的定义是:“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、音频、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,或存储在磁盘上。这种附着在各种载体上的记录统称为文献。”

从定义中不难看出文献包括4个要素。

(1) 知识和信息,即文献的内容。

(2) 物质载体,如竹简、纸张、胶片、磁盘等,它是文献的外在形式。

(3) 记录知识和信息的符号,如文字、图表、声音、图像等。

(4) 记录的方式或手段,如书写、印刷、复制、录音、录像等,即将知识和信息固化到载体上的手段。

文献是记录有知识和信息的一切载体,或称为固化在某种物质载体上的知识和信息,当文献中记录的知识传递给用户,并为用户所利用时,文献中的知识就转化为情报。因此,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。现在通常将其理解为图书、期刊等各种出版物的总和。

5. 信息、知识、情报与文献

生活中信息无处不在,无时不有,它们是知识产生的原料,这些原料经过人脑接收、

处理后,成为系统化的信息,知识由此产生。从知识管理的层面上看,知识是指可直接用于行动的信息,它使人们可以随时随地做出正确决策。

目前,我国图书情报学界对信息、知识、文献和情报的看法是:知识是对信息加工、吸收、提取、评价的结果,即系统化的信息成为知识,知识记录下来成为文献,文献经传递并加以应用成为情报,情报体现了人运用知识的能力。

1.1.2 信息的分类

信息的分类没有固定的标准,可以根据不同需求从多个角度进行划分。

1. 按信息的内容划分

按信息的内容划分,其可分为科学技术信息和社会科学信息。科学技术信息包括自然科学信息、工程技术信息等;社会科学信息包括政治信息、军事信息、经济信息、技术经济信息、社会生活信息等。

2. 按产生信息的客体的性质来划分

按产生信息的客体的性质来划分,其可分为自然信息(瞬时发生的声、光、热、电,形形色色的天气变化,缓慢的地壳运动,天体演化等)、生物信息(生物为繁衍生存而表现出来的各种形态和行为,如遗传信息、生物体内的信息交流、动物种群内的信息交流)、机器信息(自动控制系统)和社会信息(人类)。社会信息是指人与人之间交流的信息,既包括通过手势、身体、眼神所传达的非语义信息,也包括用语言、文字、图表等语义信息所传达的一切对人类社会运动变化状态的描述。

3. 按信息所依附的载体来划分

按信息所依附的载体来划分,其可分为文献信息、口头信息等。文献信息,就是以文献为载体的信息,以文字、符号、声像信息为编码信息,是经人们筛选、归纳和整理后记录下来的,它与人工符号本身没有必然的联系,但需要通过符号系统实现传递。口头信息是原始的没有形成文字、符号、声像的信息,不如文献信息易于流传。

1.1.3 文献的分类

文献是信息存在的一种重要形式,其优点是易识别、易保存、易传播,使人类精神文明千古流传。文献是信息的主要载体,大多规范化的重要信息都以文献形式出现,所以文献的分类非常重要。

1. 按照文献的加工层次分类

文献是信息的主要载体,根据对文献信息的加工层次可将文献分为零次文献、一次文献、二次文献和三次文献。

(1) 零次文献(Zeroth Literature),指未经正式发表或未形成正规载体的一种文献形式,如书信、手稿、文献草稿、会议记录、实验笔记、口头演讲等。零次文献具有客观性、针

对性、直观性强等特点,但也有传播面窄、难以保管以及信息零散和不成熟的缺点。零次文献一般是通过口头交谈、参观展览、参加报告会等途径获取,不仅在内容上有一定的价值,而且能弥补一般公开文献从信息的客观形成到公开传播之间费时甚多的弊病。

(2) 一次文献(Primary Literature),指人们直接以自己的生产、科研、社会活动等实践经验为依据生产出来的文献,也常被称为原始文献(或叫一级文献),其所记载的知识、信息比较新颖、具体、详尽,如期刊论文、专利说明书、科技报告、会议论文、学位论文、技术标准等。这些文献具有创新性、实用性和学术性等明显特征,是科技查新工作中进行文献对比分析的主要依据。

(3) 二次文献(Secondary Literature)又称检索性文献,指对一次文献按照其外部特征和内容特征进行有序化加工、整理所形成的文献形式,如目录、题录、索引、文摘等。二次文献在内容上不具有原创性,只是一次文献特征的汇总,提供检索一次文献的线索。二次文献内容相对集中、系统性强,便于管理和传播,是查新工作中检索文献所利用的主要工具。

(4) 三次文献(Tertiary Literature)又称参考性文献,指围绕某个专题,在一、二次文献的基础上,经过筛选、综合、分析和浓缩等深度加工而形成的文献,如专题评述、动态综述、进展报告以及手册、百科全书、年鉴、大全、词典等工具书。三次文献是对现有成果加以评论、综述并预测其发展趋势的文献,具有系统性、综合性、知识性和概括性等特点。在查新工作中,利用三次文献可以在短时间内了解其研究历史、发展动态及水平等,以便能更准确地掌握待查项目的技术背景,把握查新点。

总之,从零次文献、一次文献、二次文献到三次文献,是一个由分散到集中,由无序到有序,由多到精的对知识信息进行不同层次加工的过程。它们所含信息的质和量是不同的,对于改善人们的知识结构所起到的作用也不同。零次文献和一次文献是最基本的信息源,是文献信息检索和利用的主要对象;二次文献是一次文献的集中提炼和有序化,它是文献信息检索的工具;三次文献是把分散的零次文献、一次文献和二次文献,按照专题或知识的门类进行综合分析加工而成的成果,是高度浓缩的文献信息,它既是文献信息检索和利用的对象,又可作为检索文献信息的工具。

2. 按照文献的载体分类

随着信息记录与存取技术的发展,文献载体形式呈现多样化,如音像磁带、缩微胶卷、光盘等,这些非纸型文献的出现使文献的范围进一步扩大;使文献的生产和传递更加迅速;使知识、信息的存储和利用更加便捷。目前文献主要有印刷型、缩微型、电子型(机读型)、音像型4种。

(1) 印刷型文献(Paper Literature),指以手写、打印、印刷等为记录手段,将信息记载在纸张上形成的文献。它是传统的文献形式,便于阅读和流传,但存储密度小、体积大,不便于管理和长期保存。

(2) 缩微型文献(Microform),指以感光材料为载体、用缩微照相技术制成的文献复制品,如缩微胶卷、缩微平片。其特点是存储密度大、体积小,便于保存和传递,但必须借

助专门的设备才能阅读。世界上许多文献信息服务机构都将欲长期收藏的文献制成缩微品加以保存。

(3) 电子型文献(Electronic Literature), 指以数字代码方式将图、文、声、像等信息存储到磁、光、电介质上, 通过计算机或类似设备阅读使用的文献, 也称机读型文献, 如各种电子图书、电子期刊、联机数据库、网络数据库、网络新闻、光盘数据库等。其特点是信息存储量大, 出版周期短, 易更新, 传递信息迅速, 存取速度快, 可以融文本、图像、声音等多媒体信息于一体, 信息共享性好, 易复制。

(4) 音像型文献(Audio-Visual Literature), 是采用录音、录像、摄影、摄像等手段, 将声音、图像等多媒体信息记录在光学材料、磁性材料上形成的文献, 又称声像资料、视听资料、音像制品, 如音像磁带、唱片、幻灯片、激光视盘等。音像文献脱离了传统的文字记录形式, 直接记录声音和图像, 给人以直观的感觉, 又称直感型资料。其特点是形象、直观, 尤其适于记录用文字、符号难以描述的复杂信息和自然现象, 但其制作、阅读需要利用专门设备。

3. 按照文献的出版形式分类

按照文献的出版形式和内容, 文献可以分为图书、期刊、报纸、特种文献(学位论文、会议论文、专利文献、标准文献、科技报告、政府出版物、产品样本资料等)。

(1) 图书。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, 具有特定的书名和著者名, 编有国际标准书号(ISBN), 有定价并取得版权保护的出版物称为图书。

图书是以传播知识为目的, 用文字或其他信息符号记录于一定形式的材料之上的著作物; 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。其包括: 专著、教科书、词典、丛书、工具书、百科全书等。

知识窗: ISBN 号

国际标准书号(International Standard Book Number, ISBN)是国际通用的图书或独立的出版物代码。一个国际标准书号只有 1 个或 1 份相应的出版物与之对应。国际标准书号由 13 位数字组成。前 3 位数字代表图书, 中间的 9 个数字分为 3 组, 分别表示国家代码、出版社代码和书序码, 最后一个数字是校验码, 从 0 到 9 或 X。

格式举例:《科技信息检索》, ISBN 为 978-7-030-26151-9。

(2) 期刊, 也称杂志, 是由多位作者撰写的不同题材的作品构成的定期出版物。期刊有固定刊名, 以期、卷、号或年、月为序。期刊是指定期或不定期连续出版的印刷读物, 每期的内容不重复。期刊出版单位出版期刊, 必须经新闻出版总署批准, 持有国内统一连续出版物编号(ISSN)。

根据期刊的出版周期, 期刊可以分为: 旬刊、半月刊、月刊、双月刊、季刊、半年刊、年刊。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。学术类期刊按主

管单位的不同,可以分为省级、国家级、科技核心期刊(统计源期刊)、中文核心期刊(北大中文核心)、中文社会科学引文索引(CSSCI)、中国科学引文数据库(CSCD)、双核心期刊等。

知识窗: ISSN 号

国际标准连续出版物编号(International Standard Serial Number, ISSN)是根据国际标准 ISO3297 制定的连续出版物国际标准编码,是连续出版物的唯一代码标识。该编号是以 ISSN 为前缀,由 8 位数字组成。8 位数字分为前后两组,每组 4 位,中间以连接号相连。

格式举例:《中国经济问题》,ISSN 为 1000-4181。

(3) 报纸,也是连续出版物的一种,是以刊载新闻和时事评论为主的定期向公众发行的印刷出版物。其是大众传播的重要载体,具有反映和引导社会舆论的功能。根据出版周期,报纸可分为日报、早报、晚报、双日报、周报、旬报等。

(4) 学位论文,是指学生为了获得所修学位,按要求所撰写的论文。学位论文是学术论文的一种形式,有严格的格式要求,一般不公开出版。学位论文分为学士论文、硕士论文、博士论文 3 种。

(5) 会议论文,是指在会议等正式场合宣读的首次发表的论文。会议论文属于公开发表的论文,一般正式的学术交流会议都会出版会议论文集。会议论文集不是期刊,但是有的期刊为会议论文出增刊。

(6) 专利文献,是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。

(7) 标准文献,是经公认权威机构(主管机关)批准的一整套在特定范围(领域)内必须执行的规格、规则、技术要求等规范性文献,简称标准。

(8) 科技报告,是记录某一科研项目调查、实验、研究的成果或进展情况的报告,又称研究报告、报告文献。每份报告自成一册,通常载有主持单位、报告撰写者、密级、报告号、研究项目号和合同号等。其按内容可分为报告书、论文、通报、札记、技术译文、备忘录、特种出版物。

(9) 政府出版物,是指由政府机构制作出版或由政府机构编辑并授权指定出版商出版的文献。常见的政府出版物有报告、公报、通报、通讯、文件汇编、会议录、统计资料、图表、地名词典、官员名录、国家机关指南、工作手册、地图集以及传统的图书、期刊,也包括缩微、视听等其他载体的非书资料。

(10) 产品样本资料,是指厂商或贸易机构为宣传和推销其产品而印发的资料,如产品目录、产品说明书、产品总览、产品手册等。

1.2 信息检索基础

信息时代一个突出的特点是信息爆炸。人们每天都面临着来自四面八方的庞大信息量,快速准确地找出需要的信息,提高信息的利用率,掌握有效的信息检索方法是很有必要的。

1.2.1 信息检索的原理

1. 信息检索的概念

检索“Retrieval”即“查找”之意。信息检索的定义有广义和狭义之分,广义的信息检索“Information Retrieval”是指将信息按照一定的方式组织和存储起来,并根据信息用户的需要揭示、查找、传递相关信息的过程,包括了信息的存储过程和查找过程。狭义的信息检索仅指信息的查找,是指从信息集合中找出用户所需要的有关信息的过程。狭义的信息检索包括3个方面的含义:了解用户的信息需求;掌握信息检索的技术或方法;满足信息用户的需求。

国内外有关专家关于信息检索给予了不同的解释,较有代表性的观点主要有以下几种。

(1) 信息检索的范围较大。动态信息、静态信息、声频信息、视频信息及各种数值信息均属信息检索范围。如果将信息检索作为一门学科,它应该包括矩阵记数法、概率论、最优化理论、模式识别及系统分析技术等各学科领域的内容。

(2) 信息检索主要是文献检索。信息检索是从大量的文献中查找出与情报提问所指定的课题(对象)有关的文献,或者是包含用户所需事实与消息的文献的过程。这里谈到的文献,不仅指文献线索,也包括文献的片断,如章、节、段落以及与事实有关的直接情报等。

(3) 信息检索是指将信息按一定的方式组织起来,并根据用户需求找出相关信息的过程。这是指信息的存储与检索,是针对信息工作者和用户来定义的,如果仅针对用户,信息检索是指在信息集合中找出所需信息的过程。

2. 信息检索的原理

信息检索目的就是在信息用户与信息源之间充当媒介。信息检索经历了手工检索、计算机检索、网络检索和智能化检索等多个发展阶段,其信息检索的基本原理就是在对信息进行整理排序形成检索工具的基础上,按照用户的要求利用检索工具或检索系统,将用户检索提问标识(检索词)与已形成的或存储在系统中的信息的存储标识(文献特征标识、标引词)进行匹配比较,若取得一致,则为匹配,即达到了用户的检索需求,如图1.1所示。

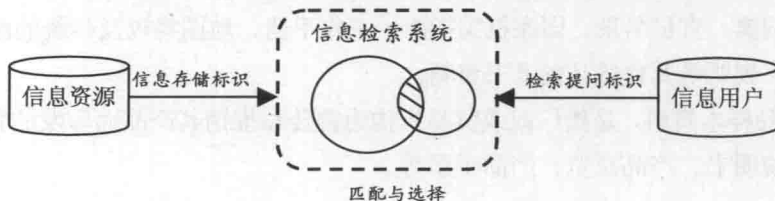


图 1.1 信息检索原理图

由信息检索原理可知,信息的存储是实现信息检索的基础。这里要存储的信息不仅包括原始文档数据,还包括图片、视频和音频等,首先要将这些原始信息进行计算机语言的转换,并将其存储在数据库中,否则无法进行机器识别。待用户根据意图输入查询请求后,检索系统根据用户的查询请求在数据库中搜索与查询相关的信息,通过一定的匹配机制计算出信息的相似度大小,并按从大到小的顺序将信息转换输出。

1.2.2 信息检索的类型

1. 按信息存储和检索的内容分类

信息检索按存储和检索的内容可划分为:文献检索、数据检索和事实检索。

(1) 文献检索。文献检索通常指的是在检索系统中检索以二次文献为对象(目录、索引、文摘)的信息,它们是文献信息的外部特征与内容特征的描述集合体。信息用户通过检索获取的是原文的“替代物”。

(2) 数据检索。数据检索指在检索系统中检索的数值型数据,如科学技术常数、各种统计数据、人口数据、气象数据、市场行情数据、企业财政数据等,即事物的绝对值和相对值的数字。检索系统提供一定的运算推导能力,例如外推、内插、填补空缺数据,甚至列出曲线图或进行各种分析等功能。信息用户可用通过检索获得的经过核实、整理的数值信息再做定量分析。

(3) 事实检索。事实检索指在检索系统中的原始文献中检索关于某一事物(事件、事实)发生的时间、地点和过程(情况)等方面的信息。它包括数值信息和系统数据信息。一般从系统中检索出所需信息后,再加以逻辑推理才能给出结论。这类信息主要是用于管理决策。

以上3种信息检索类型的主要区别在于:数据检索和事实检索是要检索出包含在文献中的信息本身,而文献检索则检索出包含所需要信息的文献即可。

2. 按照检索系统中信息的组织方式分类

信息检索按系统中信息的组织方式可划分为全文检索和多媒体检索。

(1) 全文检索。全文检索是指检索系统中存储的是整篇文章乃至整本书。检索时,用户可以根据需要从中获取有关的章、段、句、节等信息,还可以进行各种频率统计和内容分析。随着计算机容量与运算速度的增大和提高,全文检索正迅速扩大到各学科、专业。

(2) 多媒体检索。多媒体检索是对超文本检索的补充。其存储对象超出了文本范畴,融入了静、动态图像(形)以及声音等多种媒体信息。信息的存储结构从单维发展到多维,存储空间范围在不断扩大。

3. 按照信息存储的载体和实现查找的技术手段分类

信息检索按照信息存储的载体和实现查找的技术手段可划分为以下3种:

(1) 手工检索;