



工业和信息化普通高等教育
“十二五”规划教材立项项目

21世纪高等学校规划教材

信息检索 与应用

戴庆 主编

熊馨 林洁梅 副主编

21st Century University
Planned Textbooks



人民邮电出版社
POSTS & TELECOM PRESS



工业和信息化部普通高等教育
“十二五”规划教材立项项目

21世纪高等学校规划教材

信息检索 与应用

戴庆 主编

熊馨 林洁梅 副主编

21st Century University
Planned Textbooks

人民邮电出版社

北京

信息检索与应用 / 戴庆主编. — 北京: 人民邮电出版社, 2014. 10
21世纪高等学校规划教材
ISBN 978-7-115-37026-6

I. ①信… II. ①戴… 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221131号

内 容 提 要

本书共9章, 全面介绍了信息检索概论、网络资源检索、图书检索、中文数据库检索、文献管理软件的使用、外文数据库检索、特殊文献检索、事实与数据信息检索、信息分析与应用等。为了让读者能够及时地检查自己的学习效果, 把握自己的学习进度, 每章后面都附有与重点知识点相关的习题。

本书既可以作为本科院校信息检索课程的教材, 也可以作为教学、科研、工程技术人员和对信息检索感兴趣的各界人士的自学参考资料。

◆ 主 编 戴 庆

副 主 编 熊 馨 林洁梅

责任编辑 刘盛平

执行编辑 刘 佳

责任印制 焦志炜

◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号

邮编 100164 电子邮件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

北京中新伟业印刷有限公司印刷

◆ 开本: 787×1092 1/16

印张: 15.75

字数: 415千字

2014年10月第1版

2014年10月北京第1次印刷

定价: 36.00元

读者服务热线: (010) 81055256 印装质量热线: (010) 81055316

反盗版热线: (010) 81055315

前言

信息检索起源于图书馆的参考咨询和文摘索引工作。从 19 世纪下半叶开始发展,至 20 世纪 40 年代,索引和检索已成为图书馆独立的工具和服务项目。随着电子计算机问世,计算机技术逐步走进信息检索领域,并与信息检索理论紧密结合起来。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、军事和商业等各领域高速发展,得到了广泛的应用。

本书旨在培养学生信息检索、分析和展示的能力。通过本书的学习,学生能够掌握检索工具的使用,具备根据实际需求选择合适的检索系统以及制定检索方法和策略的能力。

本书主要涵盖:信息检索概论,网络资源检索,图书检索,中文数据库检索,文献管理软件的使用,外文数据库检索,特殊文献检索,事实与数据信息检索,信息分析与应用等内容。

本书由北京城市学院信息技术教研室戴庆任主编,熊馨、林洁梅任副主编。第 1 章由林洁梅、孙圆媛编写;第 2 章由郭新、王大平编写;第 3 章由蒋衡、毛宁编写;第 4 章由熊馨、毛宁编写;第 5 章由程葳、熊馨、郭新编写;第 6 章由戴庆、梁春霞、王大平编写;第 7 章由梁春霞、孟红、高书凤编写;第 8 章由高书凤、孟红、蒋衡编写;第 9 章由孙圆媛、林洁梅编写。

本书在编撰过程中参考借鉴了一些已出版的同类教材及相关文献,在此向相关单位及作者致以诚挚的谢意!

由于编者水平有限,书中难免有疏漏之处,敬请读者批评指正。

编者

2014 年 6 月

第 1 章 信息检索概论1	第 3 章 图书检索37
1.1 信息检索的基本概念.....1	3.1 图书的基础知识.....37
1.1.1 信息检索的概念.....1	3.1.1 图书的信息组织.....37
1.1.2 信息检索的分类.....2	3.1.2 图书的信息检索.....43
1.1.3 信息资源的分类.....3	3.2 图书馆的基础知识.....48
1.2 检索工具、检索途径与检索语言.....5	3.2.1 高校图书馆.....48
1.2.1 检索工具及其分类.....5	3.2.2 公共图书馆.....51
1.2.2 检索工具的选择.....7	3.3 图书馆藏目录检索.....52
1.2.3 检索途径.....7	3.3.1 OPAC 的基础知识.....52
1.2.4 检索语言.....9	3.3.2 中国国家图书馆 OPAC.....53
1.2.5 选择检索的关键词.....11	3.3.3 首都图书馆 OPAC.....57
1.3 构造检索式.....13	3.4 馆际互借——图书联合目录检索.....58
1.4 检索的方法和步骤.....16	3.4.1 CALIS.....58
1.4.1 常用的检索方法.....16	3.4.2 BALIS.....60
1.4.2 检索一般步骤.....17	3.4.3 CASHL.....63
习 题.....18	3.5 电子图书.....65
第 2 章 网络资源检索20	3.5.1 电子图书概述.....65
2.1 网络基础知识.....20	3.5.2 数字图书馆.....66
2.1.1 计算机网络基本构成.....20	3.5.3 超星数字图书馆.....66
2.1.2 网络的分类.....21	习 题.....70
2.1.3 Internet 相关概念.....22	第 4 章 中文数据库检索73
2.2 搜索引擎的分类与工作原理.....24	4.1 CNKI 期刊与学位论文检索.....73
2.2.1 搜索引擎的分类.....24	4.1.1 CNKI 期刊检索.....73
2.2.2 工作原理.....25	4.1.2 CNKI 学位论文检索.....84
2.3 搜索引擎的检索技巧.....25	4.1.3 CNKI 检索策略.....85
2.4 常用搜索引擎检索技巧.....26	4.2 CNKI E-Learning 的使用.....86
2.4.1 百度搜索引擎的使用技巧.....26	4.2.1 CNKI E-Learning 软件简介.....87
2.4.2 Google 搜索引擎的使用技巧.....31	4.2.2 CNKI E-Learning 软件使用.....88
2.5 Web 2.0 工具简介.....33	4.3 万方数据知识服务平台.....94
2.5.1 Blog 的应用.....34	4.3.1 万方数据库简介.....95
2.5.2 Wiki 的应用.....34	4.3.2 万方检索实例.....97
2.5.3 RSS 的应用.....35	4.4 论文检测.....98
习 题.....36	4.4.1 学术不端行为.....98

4.4.2 论文检测系统.....	99	5.6.5 文献样式制作.....	134
4.4.3 维普通达论文检测实例.....	100	习 题.....	135
习 题.....	104	第 6 章 外文数据库检索.....	136
第 5 章 文献管理软件的使用.....	105	6.1 国外全文数据库.....	136
5.1 什么是文献管理软件.....	105	6.1.1 EBSCOhost 数据库.....	137
5.1.1 文献管理软件的主要功能.....	105	6.1.2 Emerald 数据库.....	142
5.1.2 常用文献管理软件介绍.....	106	6.1.3 ScienceDirect 电子期刊全文数 据库.....	146
5.2 NoteExpress 基本功能介绍.....	107	6.2 国外文摘和引文数据库.....	151
5.2.1 安装与启动.....	107	6.2.1 工程索引(EI).....	151
5.2.2 工作界面.....	108	6.2.2 ISI Web of Knowledge.....	158
5.2.3 创建数据库.....	109	习 题.....	169
5.2.4 打开数据库.....	110	第 7 章 特殊文献检索.....	170
5.2.5 关闭数据库.....	110	7.1 专利文献的检索.....	170
5.2.6 创建虚拟文件夹.....	110	7.1.1 专利文献基本知识.....	170
5.3 题录导入案例.....	111	7.1.2 专利检索工具.....	174
5.3.1 手工录入.....	111	7.1.3 专利检索实例.....	176
5.3.2 批量导入.....	113	7.2 标准文献.....	178
5.3.3 在线检索导入.....	115	7.2.1 标准文献基本知识.....	178
5.3.4 导入全文和更新题录.....	116	7.2.2 标准文献的检索.....	179
5.3.5 应用 IE 插件边浏览边导入.....	118	7.2.3 标准文献检索实例.....	181
5.4 文献管理案例.....	120	7.3 学位论文检索.....	182
5.4.1 查重.....	120	7.3.1 学位论文概述.....	182
5.4.2 检索.....	120	7.3.2 学位论文检索工具.....	182
5.4.3 附件.....	122	7.3.3 学位论文检索实例.....	184
5.4.4 笔记.....	124	7.4 会议文献检索.....	187
5.4.5 统计.....	125	7.4.1 会议文献概述.....	187
5.4.6 组织.....	127	7.4.2 会议文献检索工具.....	188
5.5 论文撰写案例.....	128	7.4.3 会议文献检索实例.....	191
5.5.1 写作插件.....	128	习 题.....	192
5.5.2 插入引文.....	128	第 8 章 事实与数据信息检索.....	194
5.5.3 格式化参考文献.....	128	8.1 事实与数据检索的概述和特点.....	194
5.5.4 定位引文.....	129	8.1.1 概述.....	194
5.5.5 格式化复制.....	130	8.1.2 特点.....	195
5.5.6 插入笔记.....	130	8.2 工具书.....	196
5.6 NoteExpress 高级功能介绍.....	131	8.2.1 传统工具书的分类和特点.....	196
5.6.1 全文下载.....	131	8.2.2 参考工具书的定义、特点、作用.....	196
5.6.2 题录导出.....	132	8.2.3 参考工具书的类型及资源.....	197
5.6.3 批量编辑和替换.....	132		
5.6.4 手稿模板.....	133		

8.2.4 参考工具书使用方法	202
8.2.5 参考工具书常用检索方法	204
8.3 事实与数据信息检索工具和实例	205
习 题	212
第 9 章 信息分析与应用	214
9.1 信息检索策略的制定与实施	214
9.1.1 信息检索策略的概念	214
9.1.2 信息检索策略的制定	215
9.1.3 信息检索策略的调整	217
9.1.4 信息检索策略的实施	217
9.2 检索效果评价	218
9.2.1 检索系统评价	219
9.2.2 检索效果评价	220
9.2.3 影响检索效果的因素	221
9.3 信息的收集、鉴别与阅读	222

9.3.1 信息的收集	222
9.3.2 信息鉴别的方法	229
9.3.3 网络信息资源的鉴别与评价	231
9.3.4 信息的阅读	232
9.4 文献综述的写作	232
9.4.1 什么是文献综述	232
9.4.2 文献综述的写作方法	233
9.5 学术论文的写作	234
9.5.1 学术论文的性质	234
9.5.2 学术论文的选题	235
9.5.3 学术论文的写作与信息资源	238
9.5.4 学术论文写作基础	241
9.5.5 学术论文写作的一般步骤	243
9.5.6 学术论文的投稿	245
9.5.7 毕业论文的答辩	246
习 题	246

第1章

信息检索概论

信息社会为人们带来丰富的信息资源的同时,也带来了信息的泛滥。知道哪里可能有我们需要的信息资源,在哪里寻找到并且获取相关的信息,去芜存菁,是信息检索要做的事情。通过信息检索可以了解自己想要研究的课题是否有人做过,了解和掌握国内外已有的成果,根据检索的结果可以调整自己的研究思路避免重复的劳动和走弯路,节约了人力、物力、财力和时间。因此,具备熟练的信息检索能力是我们想获得更好的发展时所必需的能力。

1.1 信息检索的基本概念

1.1.1 信息检索的概念

对信息检索的概念,人们有不同的理解,大体可以分为广义和狭义两类。

广义的“信息检索(Information Retrieval)”,由信息的存储与检索(Information Storage and Retrieval)两个过程组成,是将信息按一定方式组织和存储到数据库中,再根据用户的需求,从存放的地方找出并提取信息的过程,如图 1-1 所示。

狭义的“信息检索(Information Searching)”是信息的检索过程那部分,即按照某种方式从现有的信息集合或数据库中查找并提取出信息的过程。

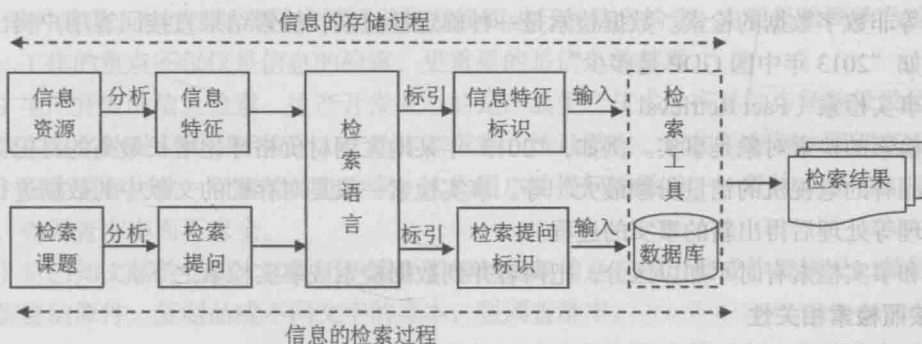


图 1-1 广义的信息检索基本原理图

其中,检索语言是沟通信息存贮和检索两个过程的桥梁。检索语言是一组有规律、能反映出信息内容及特征的标识符号,是信息处理人员和检索人员共同遵守的语言,在加工、存储及检索

信息时使用。

存储过程由专业的信息处理人员完成,从各种信息资源中搜集和整理信息,通过分析提取出能够准确全面表达信息资源内容的特征,按照检索语言的规则和要求为其设置确切的信息特征标识(即标引)并组织文献,输入检索工具,存储到数据库中。

检索过程面向以普通用户为主的信息用户。用户通过对检索课题进行分析,形成检索提问。按照检索语言的规则和要求设置检索提问标识,输入检索工具,在数据库进行查找,用户所设置的检索提问标识和数据库中存储的信息特征标识进行比较,两者相匹配的信息被检索出来,最后输出检索结果。

1.1.2 信息检索的分类

按不同的角度,信息检索有多种分类方式,下面列出常用的几种分类方法。

1. 按照信息检索的内容

(1) 文献检索 (Document Retrieval)

文献是用文字、图形、符号、声频和视频等技术手段记录人类知识的一种载体。文献包括各种图书、期刊、会议文献、科技报告、专利文献、学位论文、科技档案等各种类型的出版物,也包括用声音、图像以及其他手段记录知识的全部现代出版物。

文献检索是利用检索方式与手段,在存储文献的检索工具或数据库中,查找检索用户所需文献的过程。文献检索是最为重要、最常使用的信息检索形式,其检索对象为文献,结果是检索出与某一课题有关的若干篇论文、书刊及其来源出处、收藏地点等。例如,检索“国际贸易壁垒”方面的文献。

根据检索内容的不同,文献检索又可分为书目检索和全文检索。

① 书目检索以文献线索为检索对象,检索结果是目录、题录、文摘等与检索课题相关的一系列文献的线索,用户需要进一步查找收藏单位,才能获得原始文献。

② 全文检索是随着数据库技术、网络技术发展而发展起来的。以文献所含的全部信息为检索对象,通过检索系统存储,就可以获得整篇文章或整本图书的全部内容。

(2) 数据检索 (Data Retrieval)

数据检索的检索对象是数据,又称数值检索。它是利用参考工具书、数据库等找出包含在信息中的某一数字数据(如电话号码、银行账号、各种统计数据、参数等)和市场行情、图表、化学分子式等非数字数据的检索。数据检索是一种确定性检索,检索结果直接回答用户提出的具体问题,例如“2013年中国GDP是多少”。

(3) 事实检索 (Fact Retrieval)

事实检索的检索对象是事实。例如,“2013年某地区钢材价格环比增长最高的月份”、“本年度哪种品牌的电视机的销量份额最大”等。事实检索一般是对存贮的文献中的数据进行分析、比较、推理等处理后得出新的事实的过程。

数据和事实检索有时不加以区分,把两者并到数据检索或事实检索之下。

2. 按照检索相关性

根据检索相关性要求,信息检索分为强相关检索和弱相关检索。

① 强相关检索强调检索的准确性。用于向用户提供高度对口信息的检索,也称为特性检索。这种检索注重查准,只要检索得到的文献信息能够满足用户的需求即可,通常对于检索结果的数量多少不做要求。

② 弱相关检索强调检索的全面性,用于向用户提供系统完整信息的检索。这种检索注重查全,要求检索出一段时间期限内有关特定主题的所有信息,为了尽量避免漏检相关信息,一般对于检索的准确性要求较低。

较多的时候,用户对于检索的要求介于两者之间,既要求查找对口的信息,又希望得到所有的信息。

事实上,检索的相关性也往往成为评价检索质量优劣的指标,即查全率与查准率。查全率是指系统在进行某一检索时,检出的相关信息数与检索库中所有相关信息总数的比率;查准率(Precision Factor)是指系统在进行某一检索时,检出的相关信息数与被检出的信息总数(包括相关信息和非相关信息)的比率。

查全率=被检出的相关信息数/检索库中所有相关信息总数 $\times 100\%$

查准率=被检出的相关信息数/被检出信息总数 $\times 100\%$

依据用户使用信息的目的不同,信息检索对检索相关性有不同要求。

(1) 撰写论文的信息检索

这类检索大部分是为了写论文而要求检索较新的期刊和学位论文等文献,主要强调查准率,一般检索5~10年以内的文献。写一般的论文往往需要10篇左右的参考文献,而撰写博士论文往往需要100篇以上参考文献。

(2) 学科建设和科学研究的信息检索

为学科建设和科学研究而进行的信息检索通常要求全面而系统,强调查全率,对查准率的要求可适当放宽,允许适当的误检。不仅要求得到本学科的相关信息,而且也希望能得到相近学科和交叉学科的有关信息。这类检索非常重视事实数据,同时对时间往往也会有要求。对于研究层次低、发展快的学科,则检索的时段可以适当缩短,反之,则适当延长。

一般来说,社会科学、基础科学、高层次的科学研究都要求检索近20年的相关信息;自然科学一般要求检索近10年左右;应用科学和一般层次的研究检索近5年的相关信息就能满足要求。以下是不同的检索目的所对应的检索要求。

- 立项查新。立项查新是申报各级、各类科技计划,申请科研课题和专利时的资料查询,这类检索特别要防止漏检。

- 成果查新。成果查新是为成果鉴定和项目申报奖项而进行的信息检索,这类检索的要点是防止漏检和保证检索结果的真实性。

- 学科建设。为创建新学科或促进学科发展而进行的信息检索,主要是强调信息的完整性和系统性。工作的重点不仅仅是信息的检索,更重要的是信息的整理。

(3) 生产开发的信息检索。生产开发关心的是产品生产技术、产品的市场前景及经济效益,这类检索主要是针对专利商标信息、产品样本资料、统计数据、价格信息等,强调查准率。

(4) 回答或解决单一问题的信息检索。这类用户的提问通常单一、具体,如HN1(甲流)的防治等,强调查准率而不求全。

(5) 对已知文献的查找。这类用户需要了解的是某信息来源出处或作者情况,其目的是要得到有关信息的原件、复制品或不同文字的译本,强调查准率。

1.1.3 信息资源的分类

信息(Information)是自然界、人类社会以及思维活动中普遍存在的现象,是一切事物自身存在方式以及它们之间相互关系、相互作用等运动状态的表达。人们把各种客观存在的、可以被

人们直接或间接开发与利用的信息的集合称为信息资源。其中,文献(Literature/Document)作为信息的一种载体,用文字、图形、符号、声频、视频等技术手段记录人类知识,不仅是信息传递的主要物质形式,也是利用信息的主要手段,是一种重要的信息资源。当我们进行信息检索时,很多情况下是对文献进行检索。因此在下文中不对信息资源和文献区分。

1. 按照主要出版形式分类

① 图书。论述或介绍某一领域知识的出版物。图书往往是著者在收集大量第一手资料的基础上,经分析归纳后编写而成的。其特点是内容比较系统、全面、成熟、可靠,但出版周期较长,报道速度相对较慢。图书主要用于对大范围的问题需要获得一般性的知识或对陌生的问题需要初步了解的场合。图书又可分为三类:

- 教科书、科普读物和一般生产技术图书,属阅读性的图书;
- 辞典、手册和百科全书等,属工具性的图书;
- 含有独创性内容的专著,属原始文献。

② 期刊。期刊一般是指名称固定、开本一致的定期或不定期的连续出版物。期刊论文内容新颖,报道速度快,信息含量大,是传递科技情报、交流学术思想最基本的文献形式。据估计,期刊情报约占整个情报源的60%~70%,因此,受到科技工作者的高度重视。大多数检索工具也以期刊论文作为报道的主要对象。对某一问题需要深入了解时,较普遍的办法是查阅期刊论文。

③ 会议文献。这是指在国际或国内重要的学术或专业性会议上发表的论文。会议文献学术性强,往往代表着某一领域内的最新成就,反映了国内外的科技发展水平和趋势,是获得最新情报的一个重要来源。会议文献可分为会前文献(如会议日程预报和会议论文预印本)以及会后文献(如各种会议录)。会后文献是主要的会议文献。

④ 科技报告。科技报告是指国家政府部门或科研生产单位关于某项研究成果的总结报告,或是研究过程中的阶段进展报告。报告的出版特点是各篇单独成册,统一编号,由主管机构连续出版。在内容方面,报告比期刊论文等专深、详尽、可靠。

⑤ 专利文献。专利文献主要由专利说明书构成。所谓专利说明书是指专利申请人向专利局递交的有关发明目的、构成和效果的技术文件。它经专利局审核后,向全世界出版发行。专利说明书的内容比较具体,有的还有附图,通过它可以了解该项专利的主要技术内容。由于只有符合新颖性、创造性和实用性的发明创造才能获得专利权,所以专利说明书对于工程技术人员,特别是产品工艺设计人员来说,是一种切合实际、启迪思维的重要情报源。

⑥ 标准文献。指标准化工作的文件。其中主要为工业产品和工程建设的质量、规格和检验方法等的技术规定文件。

⑦ 学位论文。学位论文是指为申请硕士、博士等学位而提交的学术论文。学位论文的质量参差不齐,但都是就某一专题进行研究而做的总结,多数有一定的独创性。学位论文是非卖品,除极少数以科技报告、期刊论文的形式发表外,一般不出版。

⑧ 产品技术资料。指产品目录、产品样本和产品说明书一类的厂商产品宣传和使用资料。产品样本通常对定型产品的性能、构造、用途、用法和操作规程等做具体说明,内容成熟,数据可靠,有的有外观照片和结构图,可直接用于产品的设计制造中参考。

⑨ 技术档案。指科研生产活动中形成的,有具体事物的技术文件、图纸、图表、照片和原始记录等。详细内容包括任务书、协议书、技术指标、审批文件、研究计划、方案大纲、技术措施、调查材料、设计资料、试验和工艺记录等。这些材料是科研及生产工作中用以积累经验、吸取教

训的重要文献。技术档案一般为内部使用,不公开出版发行,有些有密级限制,因此在参考文献和检索工具中极少引用。

⑩ 政府出版物。指各国政府部门及其设立的专门机构发表的文献。政府出版物的内容十分广泛,既有科学技术方面的,也有社会经济方面的。就文献的性质而言,政府出版物可分为行政性文件(如国会记录、政府法令、方针政策、规章制度以及调查统计资料等)和科学技术文献两部分。我国政府发表的《科学技术白皮书》就是一种科技类政府出版物。

2. 按照加工程度分类

依内容性质和加工程度的不同,可分为以下四个级别。

(1) 零次文献。未经出版发行的文献,包括手稿、个人通信、原始记录等。

(2) 一次文献。首次出版的各种文献,也称原始文献。如期刊论文、科技报告、会议论文、专利说明书等。一次文献是以科研生产活动的第一手成果为依据而创作的文献,内容丰富,参考价值大,是我们检索和利用的主要对象。

(3) 二次文献。报道和查找一次文献的检索书刊,如各种目录、题录和文摘等。二次文献是图书情报工作者在大量收集原始文献的基础上,经过分析、归纳、重组后出版的。二次文献是一次文献的集约化、有序化的再次出版,是储藏、利用一次文献的主要的、科学的途径。

(4) 三次文献。利用二次文献提供的线索,选用大量一次文献的内容,经综合、分析和评述再度出版的文献,如各种述评、进展报告、动态综述、手册、年鉴和百科全书等。我们使用的各种教科书也属三次文献。

零次文献由于没有进入出版、发行和流通这些渠道,收集利用十分困难,一般不能作为我们利用的文献类型。而后三种文献是一个从分散的原始文献到系统化、密集化的过程。一般说,一次文献是基础,是检索利用的对象;二次文献是检索一次文献的一种检索工具;三次文献是一次文献内容的高度浓缩,也是一种重要情报源,同样也是检索利用的对象。

【举例】研究的文稿如果希望以最快的形式发表,有哪些途径可以选择?

各种文献类型在出版时间上是有先后的。其中会议文献、科技报告可以比较快地发表,专利也可以比较快地申请下来。期刊品种多,容量大,速度也较快,是多数论文的发表渠道,所形成的文献是一次文献。目录、题录和文摘,略迟于专利、报告、会议和期刊,是对原始文献的主要内容做客观报道,所形成的文献是二次文献。而综述、图书和百科全书则需取大量原始文献的精华,或做评论,或汇编成册,所需时间更长,所形成的文献是三次文献。以上的过程从出版形式上说是文献周期,这种文献周期又称为文献链。

1.2 检索工具、检索途径与检索语言

1.2.1 检索工具及其分类

检索工具是人们用来存储、报道和查找各类信息的工具。是对信息进行搜集、整理、分析、加工和组织而形成的产物。检索工具有两个基本功能:一个是存储功能,存储原始信息或者原始信息的替代品,通过一定的检索语言、描述信息的内容特征和形式特征,按照一定的分类体系或者方法组织起来,从而完成信息存储过程;另一个功能是检索功能,即提供各种检索手段,使人们可以按照既定的检索方法,从中找到所需的原始信息或者原始信息的替代品。

按照不同的角度,检索工具可以分为多种类型。

1. 按照记录信息的技术划分

(1) 印刷型检索工具,其主要的检索工具是书本型、卡片型的信息系统,即目录、索引、文摘和各种工具书,由人通过手工的方式完成检索的过程。

(2) 机械型检索工具,主要指利用力学、光学、电子学等手段帮助检索的工具,如机检穿孔卡片、光电检索系统、计算机检索系统等。

随着信息时代的到来,计算机检索已成为主流,面向计算机和网络的各种数据库检索系统以及搜索引擎这类检索工具的利用程度越来越高,计算机检索已广泛应用于现代信息检索过程,是本书的讲述重点。

2. 按照所收信息的学科范围划分

检索工具中所收信息的学科内容有综合性、专业性之分。

(1) 综合性检索系统。综合性检索系统收录范围和涉及学科广、信息类型和语种多,因而适应面较广,是科研工作最常用的检索系统。世界著名的综合性检索系统有美国的《工程索引(EI)》《科学引文索引(SCI)》、英国的《科学文摘》等,我国最为常用的有“中国知网”“万方数据知识服务平台”“维普信息资源系统”等。

(2) 专业性检索系统。专业性检索系统的收录范围限于某一学科领域,适应于检索专业信息。如《中国石油文摘》《中国化学化工文摘》等,都是专业性检索系统。

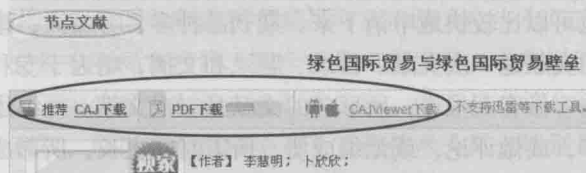
3. 按记录信息的完整程度划分

(1) 全文检索工具

全文检索工具主要指全文数据库,它可以在全文中的字、词和句中进行检索,并且为检索结果提供全文,如图1-2(a)、图1-2(b)所示。全文数据库不仅包括一次文献,还可以包括三次文献,例如“中国知网”《中国统计年鉴》全文数据库。



(a) 进行全文检索



(b) 提供全文下载

图 1-2 全文检索工具

(2) 二次检索工具

① 目录型检索工具

目录一般以整本的图书、期刊等作为报道单元,描述比较简单。每一个条目的著录项有:书(刊)名、卷(册)数、作者、出版年月、出版地和书(刊)收藏情况等,主要用于查找出版物的出版或收藏单位,例如《全国总书目》。

② 题录型检索工具

题录是将文献的篇目按照一定的排检方法编排而成的,供人们查找篇目出处的工具。题录是在目录的基础上发展起来的,以出版物中的“篇”作为著录单元的检索系统,一篇文献作为一条款目,著录项至少包括篇名、责任者和文献的出处,例如《中文社科报刊篇名数据库》。

③ 文摘型检索工具

文摘是将论文或专著的内容加以浓缩,以精练的语言把文献信息的重要内容、学术观点、数据及结构准确地摘录下来,并按一定的著录规则与排列方式编排起来,其实质就是题录加上内容摘要。这种工具为题录型检索工具的进步,人们利用它既可以检索相关信息的线索,同时还可以通过信息的摘要大致了解信息的内容,从而判断是否需要阅读原文。例如《中国生物学文摘》、美国《化学文摘》。

④ 索引型检索工具

索引一般是附在专著或年鉴、百科全书等工具书之后以及收录内容较多的二次文献之后,按主题词、人名、地名、事件、概念等内容要项编排,是查找隐含在文章中所需情报,进行微观检索的有用工具。按其内容可分为主题索引、分类索引、关键词索引及引文索引。最通用的索引是主题索引和著者索引,例如《全国报刊索引》的索引。

1.2.2 检索工具的选择

检索工具的选择应该根据检索者对信息的要求及条件来确定。一般来说,要考虑以下因素。

① 根据检索的目的选择检索工具。例如:如果需要了解国内外科技文献的发展概况、动态和水平,以及查找以往各年的早期文献,可考虑使用期刊式检索工具;要进行设计、研发新产品,需要了解新技术及新动态,应先对专利或科技报告等文献进行检索;要查找某个学科的沿革和发展概况,可以考虑检索图书;检索新信息,最好是用网络搜索引擎。为了避免检索工具在编辑出版过程中的滞后性,还应该在必要时补充查找若干主要相关期刊的现刊,以防止漏检。

② 根据要检索的信息所属的专业领域选择检索工具。通常要以专业性检索工具为主,再配合以综合型检索工具。如果要检索多个专业的文献,则要首先选用综合性检索工具。

③ 根据要检索的文献的语言选择中文或外文检索工具。

④ 选择更新速度快的检索工具。更新越快,就越能查找到最新的信息,例如 EI 数据库有印刷版、光盘版、网络版,国内收藏的 EI 数据库通常网络版更新最快。

另外,选择检索工具时还可以考虑检索工具所覆盖的学科范围、信息类型、检索的速度、数量和方便程度、经济能力等因素。

一般来说,选择检索工具时,先考虑选择国内的检索工具,后选择国外的;先选择本地图书馆后选择其他图书馆;先查文摘型检索工具,必要时再查其他工具。通常还应需要根据信息内容互补原则采用多种工具构成信息完整的检索工具体系。

1.2.3 检索途径

信息被存储的时候,设置了多个项目或字段,例如题名(包括书名、刊名、篇名等)、关键词、作者、摘要等,检索时这些项目或字段就是检索入口,又被称为检索途径。

每个检索工具都提供了自己的检索途径,主要有两大类途径,即内部和内容特征途径及外部的形式特征途径。

- 内部的内容特征途径,包括主题途径和分类途径;
- 外部的形式特征途径,包括作者途径、机构途径、号码途径和引文途径等。

1. 主题途径

主题途径是根据文献的主题特征,利用各类主题目录和索引进行检索的途径。

主题是一组具有共性事物的总称,用来表达文献所论述和研究的具体对象和问题,是文献的

“中心内容”。每种文献都包含着若干主题,研究或阐述一个或多个问题。

在主题目录和主题索引中,文献按表现其内容特征的主题词组织。检索时,只要找到所查主题词,就可查得相关文献。这是一种最常用、最重要的检索途径。

主题词

主题词是表达主题概念的词汇。

狭义的主题词仅指叙词,叙词指收录在《汉语主题词表》等词表中可用于表达文献主题概念的,经过规范化处理的词或词组,图书情报界通常把叙词称为主题词。

广义主题词可以分为规范词汇和自由词汇,包括关键词、标题词、叙词等。

① 规范词汇是从待检数据库的叙词表或主题词表中选取的规范化的、能表达文献中心内容的词或词组,词表是数据库和检索者必须共同遵循使用的检索语言。表达同一主题概念只能用一个主题词来表达。主题途径适合特性检索,查准率高,但是查全率较低,对非专业人士来说难以使用。

② 自由词汇是论文主题中新技术、新学科中新产生的尚未被主题词表收录的名词术语或在叙词表中找不到的词。用自由词在篇名、文摘甚至全文中查找信息有时比用主题词更有优越性,更直接简明,是科技人员易为接受、较为常用的一种方法,特别适合不懂得主题词的一般用户。在检索网络信息时一般使用自由词作为检索词。

2. 分类途径

分类检索系统中用分类法来表达各种信息资源的概念,将各种概念按照学科、专业性质进行分类和系统排列。分类途径就是一种在检索时,根据文献的内容特征,利用分类目录或分类索引查找文献的途径。分类法具有很好的层次性和系统性,分类目录可以向用户展示一个科学分类系统,用户可以按照知识关系查找,便于从学科体系的角度获得较系统的文献线索,传统的文献组织大多采用这种方法。

分类检索途径在我国具有悠久的历史,比较有影响的有《中国图书馆分类法》《国际十进制分类法》《杜威十进分类法》等。许多目录大多以分类方法编排。利用这一途径检索文献,首先要明确课题的学科属性、分类等级,获得相应的分类号,然后逐类查找。

3. 著者途径

根据文献的外部特征,利用著者(个人或单位著者)目录和著者索引进行检索的途径。国外比较重视著者途径的利用,许多检索工具和著作都把著者索引作为最基本的辅助索引。它是按著者的姓名字顺,将有关文献排序而成。以著者为线索可以系统、连续地掌握他们的研究水平和研究方向。

4. 其他途径

包括篇名途径、序号途径和分子式途径等途径。

根据文献的序号特征,利用其序号索引进行检索的途径。许多文献具有唯一性或一定的序号,如专利号、文摘号、国际标准图书编号、电子元件型号等。根据各种序号编制成了不同的序号索引,在已知序号的前提下,利用序号途径能查到所需文献,满足特性检索的需要。利用序号途径,需对序号的编码规则和排检方法有一定的了解;往往可以从序号判断文献的种类、出版的年份等,有助于文献检索的进行。序号途径一般作为一种辅助检索途径。

文献检索主要根据文献的外部特征和内部特征的描述,利用各种途径获得所需的相关文献。除了以上所讲的四种途径以外,还有引文途径、分子式途径、化学物质途径等,只有根据课题的需要,选用相应的检索途径,才能获得相关的文献。

【举例】论文和图书扉页中包含的检索途径信息。

论文中常见的检索途径包括：全文、主题、篇名、关键词、作者、摘要、中图分类号、参考文献、文献来源等，见图 1-3。

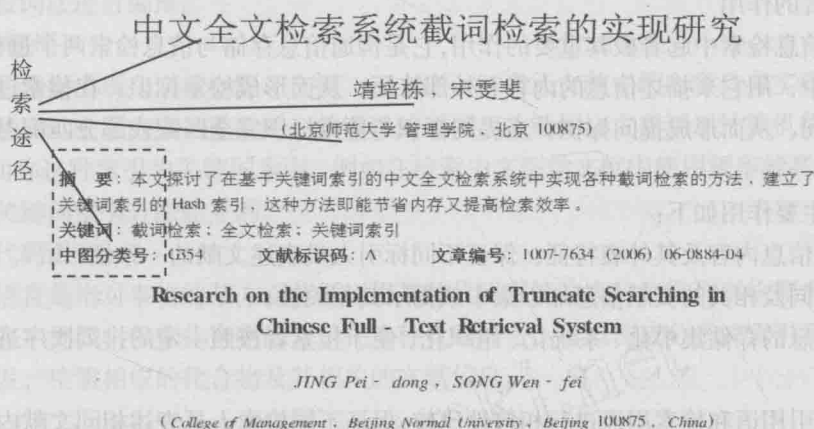


图 1-3 论文中常见的检索途径

图书的检索途径通常包括书名、主题词、ISBN、图书分类号、出版社等。打开一本书的版权页（扉页的背面），例如《贝叶斯网络在智能信息处理中的应用》一书，见图 1-4，可以看到书名、作者、出版社信息，还可以看到本书的主题词“贝叶斯推断—应用—人工智能—信息处理”、分类号码“TP18”以及 ISBN 号“978-7-118-07813-8”，这些都是检索图书时常见的检索途径。

【举例】以《美国医学索引（Index Medicus，缩写为 IM）》为例，了解检索工具的类型及检索途径。

《美国医学索引》，是目前世界上最常用的也是最权威的一种医学文献检索工具。它收录了世界各国主要的生物医学及相关学科的期刊 3300 余种，年文献报道量约为 40 万条，它是题录型检索工具。

IM 检索途径有两种：主题途径和著者途径。IM 为月刊，每年一卷，每卷随第一期出版《医学主题词表（Medical Subject Headings，缩写为 MeSH）》和《List of Journal Indexed》。IM 题录正文是按 MeSH 的主题词编排，后附有著者索引。IM 的主题词是有严格的规定的，必须依据《MeSH》中的主题词。每篇题录依其主题内容归入相应的主题词下。

1.2.4 检索语言

检索语言是和检索途径相对应的一组有规律、能反映出信息内容及特征的标识，在加工、存储及检索信息时使用，是信息处理人员和检索人员共同遵守的语言。

1. 检索语言及其作用

(1) 检索语言的概念

检索语言是应文献信息的加工、存储和检索的共同需要而编制的专门语言，是表达一系列概括文献信息内容和检索课题内容的概念及其相互关系的一种概念标识系统。简言之，检索语

图书在版编目(CIP)数据

贝叶斯网络在智能信息处理中的应用/肖秦琨,高嵩著. —北京:国防工业出版社,2012.1
ISBN 978-7-118-07813-8

I. ①贝... II. ①肖... ②高... III. ①贝叶斯推断—应用—人工智能—信息处理 IV. ①TP18

中国版本图书馆 CIP 数据核字(2011)第 267588 号

图 1-4 图书常见的检索途径

言是用来描述信息源特征和进行检索的人工语言,可分为规范化语言和非规范化语言(自然语言)两类。

(2) 检索语言的作用

检索语言在信息检索中起着极其重要的作用,它是沟通信息存储与信息检索两个过程的桥梁。在信息存储过程中,用它来描述信息的内容和外部特征,从而形成检索标识;在检索过程中,用它来描述检索提问,从而形成提问标识;当提问标识与检索标识完全匹配或部分匹配时,结果即为命中文献。

检索语言的主要作用如下:

- ① 标引文献信息内容及其外表特征,保证不同标引人员表述文献的一致性;
- ② 对内容相同及相关的文献信息加以集中或揭示其相关性;
- ③ 使文献信息的存储集中化、系统化、组织化,便于检索者按照一定的排列次序进行有序化检索;
- ④ 便于将标引用语和检索用语进行相符性比较,保证不同检索人员表述相同文献内容的一致性,以及检索人员与标引人员对相同文献内容表述的一致性;
- ⑤ 保证检索者按不同需要检索文献时,都能获得最高查全率和查准率。

2. 检索语言的类型

下面介绍几种常用的检索语言。

(1) 分类语言

分类语言是指以数字、字母或字母与数字结合作为基本字符,采用字符直接连接并以圆点(或其他符号)作为分隔符的书写法,以基本类目作为基本词汇,以类目的从属关系来表达复杂概念的一类检索语言。以知识属性来描述和表达信息内容的信息处理方法称为分类法。著名的分类法有《国际十进分类法》《美国国会图书馆图书分类法》《国际专利分类表》、《中国图书馆图书分类法》等。

(2) 主题语言

主题语言是指以自然语言的字符为字符,以名词术语为基本词汇,用一组名词术语作为检索标识的一类检索语言。以主题语言来描述和表达信息内容的信息处理方法称为主题法。主题语言又可分为标题词、元词、叙词和关键词等。

• 标题词

标题词是指从自然语言中选取并经过规范化处理,表示事物概念的词、词组或短语。标题词是主题语言系统中最早的一种类型,它通过主标题词和副标题词固定组配来构成检索标识,只能选用“定型”标题词进行标引和检索,反映文献主题概念必然受到限制,不适应时代发展的需要,目前已较少使用。

• 元词

元词又称单元词,是指能够用以描述信息所论及主题的最小、最基本的词汇单位。经过规范化的能表达信息主题的元词集合构成元词语言。元词法是通过若干单元词的组配来表达复杂的主题概念的方法。元词语言多用于机械检索,适于用简单的标识和检索手段(如穿孔卡片等)来标识信息。

• 叙词

叙词是指以概念为基础,经过规范化和优选处理的具有组配功能并能显示词间语义关系的动态性的词或词组。一般来讲,选做的叙词具有概念性、描述性、组配性。经过规范化处理后,还