



高等教育“十一五”规划教材

高职高专专业基础课教材系列

# 化工信息检索

冯西宁 李 春 主编



科学出版社  
[www.sciencep.com](http://www.sciencep.com)

高等教育“十一五”规划教材

高职高专专业基础课教材系列

# 化工信息检索

冯西宁 李 春 主编

科学出版社

北 京

## 内 容 简 介

本书从化学化工类专业的特殊性出发, 兼顾高职高专院校应用性人才的培养需求, 全书贯彻“轻理论, 重实践”的指导思想, 着重介绍化学、化工、材料等专业的电子信息资源的检索, 突出实践性、实用性和新颖性。全书分为十章主要介绍相关数据库的检索技能及利用方法, 并附有大量图例、实例和习题。

本书适合化学、化工、材料、药学和冶金类专业学生、工程技术人员和相关行业的管理人员使用。

### 图书在版编目(CIP)数据

化工信息检索/冯西宁, 李春主编. —北京: 科学出版社, 2010.2  
(高等教育“十一五”规划教材·高职高专专业基础课教材系列)  
ISBN 978-7-03-026660-6

I. ①化… II. ①冯… ②李… III. ①化学-情报检索-高等学校: 技术学校-教材 ②化学工业-情报检索-高等学校: 技术学校-教材  
IV. ①G252.7

中国版本图书馆 CIP 数据核字 (2010) 第 019477 号

责任编辑: 沈力匀 周 恢 / 责任校对: 柏连海  
责任印制: 吕春珉 / 封面设计: 东方人华平面设计部

科 学 出 版 社 出 版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

双 青 印 刷 厂 印 刷

科学出版社发行 各地新华书店经销

\*

2010 年 2 月第 一 版

开本: 787×1092 1/16

2010 年 2 月第一次印刷

印张: 17 1/4

印数: 1—3 000

字数: 406 000

定价: 26.00 元

(如有印装质量问题, 我社负责调换〈环伟〉)

销售部电话: 010-62134988 编辑部电话: 010-62135235 (VP04)

版权所有, 侵权必究

举报电话: 010-64030229; 010-64034315; 13501151303

高等教育“十一五”规划教材  
高职高专应用化工类专业系列教材  
编写委员会

主 任 李奠础

副主任 张 歧 林 峰 郑光洪 龚盛昭 葛 虹

委 员 (按姓氏笔画排序)

|     |     |     |     |     |
|-----|-----|-----|-----|-----|
| 丁文婕 | 丁 颖 | 于海军 | 王玉环 | 王爱军 |
| 尹美娟 | 白嘉玲 | 冯西宁 | 刘红波 | 刘 宏 |
| 安红莹 | 杨东洁 | 杨丽芳 | 李文典 | 李 春 |
| 吴雨龙 | 张小华 | 张 钧 | 张桃先 | 陈咏梅 |
| 李 景 | 李 勤 | 吾国强 | 吴 卫 | 吴丽璇 |
| 陈瑞珍 | 郑孝英 | 赵 宁 | 胡 伟 | 胡智华 |
| 洪 亮 | 徐 玲 | 高 虹 | 高洪潮 | 高瑞英 |
| 郭会灿 | 郭建民 | 黄健光 | 彭建兵 | 董 利 |
| 韩文爱 | 税永红 | 薄新党 |     |     |



## 前 言

在当今科学技术日新月异,互联网的使用日益普及的大背景下,掌握各种信息资源数据库的使用方法,已经成为大学生必须具备的重要技能。本书根据应用化工类专业人才培养目标,从化学化工类专业的特殊性出发,兼顾高职高专院校应用性人才的培养需求,全书贯彻“轻理论,重实践”的指导思想,力求反映信息检索领域的最新发展,做到理论与实践紧密结合,内容少而精。着重介绍化学、化工、材料等专业的电子信息资源的检索,突出实践性、实用性和新颖性。全书共分为十章,结合当前文献检索系统的最新发展,重点介绍相关数据库的检索技能及利用方法,并附有大量图例、实例和习题。

本书不仅适用于化学、化工、材料、药学和冶金类专业学生使用,还适用于工程技术人员和相关行业的管理人员。

全书编写大纲由成都纺织高等专科学校冯西宁拟定,并编写了第一章;成都纺织高等专科学校李春编写第二章、第七章、第九章;成都纺织高等专科学校魏玉君编写第三章、第四章;石家庄职业技术学院郝卓莉编写了第五章、第六章;成都工商职业技术学院邱梅、河北医药化工职业技术学院孙博雅编写了第八章、第十章。魏玉君对全书的图片进行了修订,冯西宁、李春对书稿全部内容进行了统稿和修订。

在本书的编写过程中,直接或间接参考、借鉴了国内外相关专著、教材、网站等的有关观点及信息,在此不一一注明,谨向有关单位、作者致以诚挚的谢意。

由于编者水平有限,书中难免存在不足之处,敬请广大读者及各位同行不吝赐教。

# 目 录

## 前言

|                               |     |
|-------------------------------|-----|
| 第一章 信息检索基础知识 .....            | 1   |
| 第一节 信息、知识、情报 .....            | 1   |
| 第二节 信息资源及类型 .....             | 4   |
| 第三节 信息检索基本原理 .....            | 7   |
| 第二章 化学化工类中文数据库检索 .....        | 27  |
| 第一节 CNKI 数据库 .....            | 27  |
| 第二节 万方数据库 .....               | 43  |
| 第三节 维普数据库 .....               | 54  |
| 第四节 超星数字图书馆 .....             | 71  |
| 第三章 化学化工类外文数据库检索 .....        | 83  |
| 第一节 美国《工程索引》(EI) 数据库 .....    | 84  |
| 第二节 美国《科学引文索引》(SCI) 数据库 ..... | 90  |
| 第三节 美国《化学文献》(CA) 数据库 .....    | 99  |
| 第四章 专利信息数据库检索 .....           | 125 |
| 第一节 专利文献简介 .....              | 126 |
| 第二节 中国专利信息数据库检索 .....         | 130 |
| 第三节 国外专利数据库检索 .....           | 133 |
| 第五章 科技报告数据库检索 .....           | 156 |
| 第一节 科技报告基础知识 .....            | 157 |
| 第二节 科技报告的检索工具 .....           | 160 |
| 第六章 学位论文数据库检索 .....           | 173 |
| 第一节 学位论文简介 .....              | 174 |
| 第二节 学位论文数据库的手工检索 .....        | 179 |
| 第三节 学位论文数据库的网上检索 .....        | 180 |
| 第七章 会议信息数据库检索 .....           | 193 |
| 第一节 学术会议与会议文献简介 .....         | 194 |
| 第二节 会议文献检索 .....              | 199 |
| 第八章 标准信息检索 .....              | 206 |
| 第一节 标准文献简介 .....              | 206 |
| 第二节 中国标准文献检索 .....            | 213 |
| 第三节 国际标准文献 .....              | 233 |

|                       |     |
|-----------------------|-----|
| 第九章 科技档案检索 .....      | 242 |
| 第一节 科技档案简介 .....      | 242 |
| 第二节 科技档案检索 .....      | 249 |
| 第十章 其他 .....          | 255 |
| 第一节 谷歌 (Google) ..... | 255 |
| 第二节 百度 (Baidu) .....  | 261 |
| 主要参考文献 .....          | 268 |

# 第一章 信息检索基础知识



## 学习目标

- (1) 增强学生情报意识, 提高获取文献信息能力, 培养学生自学能力和独立科研能力。
- (2) 了解文献、情报、信息的概念及它们之间的联系。
- (3) 了解信息源及其特征。



## 必备知识

计算机基础知识, 能使用 Internet 进行简单的网络检索。



## 选修知识

利用印刷型检索工具进行信息检索。



## 案例导入

例如, 查找有关甲醛的标准红外光谱图。

通常要进行查阅标准红外光谱, 可以从图书馆借阅 Sadler 标准图谱, 目前 Sadtler 红外光谱图数据库有部分免费图谱可供查阅, 如 <http://www.bio-rad.com/ir.html> 可进行相关链接。在 <http://www.360yiqi.com> 中也可进行图谱的在线检索。也可以同时利用 Google 或 Baidu 等搜索引擎在网络上进行搜索一些实验室网站, 在这些实验室的资源网站上有相关的一些图谱。



## 课前思考题

如何利用 Google 和 Baidu 查找关于“喜树碱”的相关信息。

## 第一节 信息、知识、情报

### 一、信息的定义、特征和分类

信息的概念十分广泛, 目前学术界并没有统一的定义, 不同学者从不同角度对信息做出各种定义。为了便于理解, 可以将信息定义为: 信息是用文字、数据或信号等形式



通过一定的传递和处理来表现各种相互联系的客观事物在运动变化中所具有特征内容的总称。因而可以这样认为,信息是事物存在的方式、形态和运动规律的表征,是事物具有的一种普遍属性,它与事物同在,存在于整个自然界和人类社会。

从信息内容及存在形式上看,信息所具有的基本特征可归结为以下四方面:

### 1. 信息具有客观性

世间一切事物都在运动中,都具有一定的运动状态和状态方式的改变,因而一切事物随时都在产生信息,即信息的产生源于事物,是事物的普遍属性,是客观存在的,是不以人们的意志为转移的,它可以被感知,被处理和存储,被传递和利用。

### 2. 信息具有时效性

信息具有较强的时效性。信息的时效性是指如果信息不能及时反映事物存在的方式和运动状态,信息就会失去效用。客观事物不断变化,信息的时效性也是不断变化的,如时效性很强的天气预报、股票信息、交易信息、科学信息等,如不能适时使用最新信息,信息就会失去原有的价值。

### 3. 信息具有传递性

信息的获取必须依赖于信息的传递。事物在运动过程中以及形态改变上所展现出来的特征是事物属性的再现,被人们认知后,就构成了信息的实质内容,并依附于一定的载体传递后,才能被接受和运用。由于信息的传递性,使信息有可能在短时间内广泛扩散。

### 4. 信息具有共享性

信息能够共享是区别信息不同于物质和能量的最主要特征,即同一内容的信息在同一时间、同一地域可以被两个以上的用户分享,其分享的信息量不会因分享用户的多少而受影响,原有的信息量也不会因此而损失或减少。对信息的收集、整理、加工、存储主要是为了使信息能够共享。另外,信息的共享性并不排除某些特殊信息的独占性和利用的有偿性,如军事、商业、专利信息。

信息的类型可从不同的角度划分。按存在的状态可分为瞬时信息和保留信息;按形成的领域可分为自然信息和社会信息;按表现的形式可分为文字信息、图像信息、语音信息等。

信息本身不是实体,而是一种对事件和事物的抽象反映,必须借助于一定的载体才能表现、传递和利用。载体是信息得以保存的物质实体。从古代的甲骨、金石、锦帛、竹简到现在的纸张、感光材料、磁性材料,信息载体和存储技术已经数次发生质的飞跃,为人类存储、检索和利用信息提供了极大的方便。

## 二、知识的定义、特征及类型

知识是人类在认识和改造世界的社会实践中获得的对事物本质认识的成果和结晶,

即人类通过有目的、有区别、有选择地利用信息,对自然界、人类社会及思维方式与运动规律的认识、分析与掌握,并通过人的大脑进行思维整合使信息系统化而构成知识,知识是人类在改造客观世界实际中所获得的认识和经验的总和,是信息的一部分,是一种特定的人类信息。因此,知识仅存在于人类社会。知识依存于信息,信息是知识的载体,知识是信息的内核。

知识的特征主要有以下六个方面:

(1) 意识性。知识是一种观念形态的东西,只有通过人类的大脑才能认识它、产生它、利用它。

(2) 信息性。信息是生产知识的原料,知识是经人类认识、理解并经思维重新整合后的系统化信息,知识是信息中的一部分。

(3) 实践性。实践是产生知识的基础,也是检验知识的标准,知识又对实践具有重大的指导作用。

(4) 规律性。人们在实践中对事物的认识,是一个无限的过程,人们在这种无限过程中所获得的知识从一定的层面上揭示了事物及其运动过程的规律性。

(5) 继承性。每一次新知识的产生,既是原有知识的继承利用、深化与发展,又是下一次知识更新的基础和前提。

(6) 渗透性。随着人类认识世界的不断深化,各种门类的知识可以相互渗透,构成知识的网状结构。

根据经济合作与发展组织(OECD)出版的《以知识为基础的经济》报告,人类现有的知识可分为四大类:

(1) Know what (知道是什么)——关于事实方面的知识。

(2) Know why (知道为什么)——关于自然原理和规律方面的知识。

(3) Know how (知道怎么做)——关于技能或能力方面的知识。

(4) Know where (知道归属谁)——关于产权归属的知识。

获取知识的途径有以下两方面:

(1) 直接来源于产生信息的客观事物。

(2) 通过信息载体或媒介。当然,获取的信息能否转化为知识,转化是否充分、完整,则取决于主体的认知能力。

### 三、情报的定义、特征

情报一词从字面上理解就是关于情况的报告。人们往往从军事上的意义来理解它。其实情报一词是外来语,它来自日语,有两种意义,一是指军事情报,相当于英语中的 Intelligence;二是指英语的 Information 一词。现在普遍使用的是后一种意义。情报就是人们在特定时间内为一定目的而传递的有特殊效应的知识或信息,它是一种普遍存在的社会现象。

情报有三方面的特征:

(1) 知识性。情报的知识性是指情报的本质是知识。没有一定的知识内容,任何东西都不能成为情报。

(2) 传递性。情报的传递性是指无论多么重要的知识,人们不知道其存在就不能成为情报。要让知识变成情报就必须经过传递运动。

(3) 效用性。情报的效用性是指运动的知识也不都是情报。只有那些能满足特定需要的运动知识才可称为情报。

#### 四、信息、知识、情报三者的关系

信息与知识的关系:知识可以说是建立在信息的基础上,经过加工与编码后创造出来的新信息。信息是知识的重要组成部分,但不是全部,只有提高、深化、系统的信息才能称做知识;信息能否转化为知识,转化的程度取决于信息接受方的认知能力。

信息与情报的关系:信息可以成为情报,但是一般要经过选择、综合、分析研究等加工过程;在信息或知识的海洋里,变化、流动、最活跃、被激活了的那一部分就是情报。

信息、知识、情报的逻辑关系为:信息>知识>情报。

## 第二节 信息资源及类型

### 一、信息资源概论

#### (一) 信息资源的定义

信息资源是信息与资源两个概念整合而衍生出的新概念。如前所述,信息是事物的一种普遍属性。资源,一般而言,是自然界及人类社会中一切对人类有用的资产和财物。结合资源概念可以这样来描述信息资源:信息并非都能成为资源,只有满足一定条件的信息才能成为信息资源,即信息资源是信息世界中对人类有价值的那一部分信息,是附加了人类劳动的、可供人类利用的信息。这是通常所指的广义的信息资源,是指信息及其相关因素的集合。因此,构成信息资源的基本要素是:信息、人、符号、载体。信息是组成信息资源的原料,人是信息资源的生产者和利用者,符号是生产信息资源的媒介和手段,载体是存储和利用信息资源的物质形式。信息中的载体信息和主体信息是信息资源的最基本的组成部分。

#### (二) 信息资源的特性

相对于其他非资源型信息,信息资源有许多特殊性。

(1) 共享性。共享性是信息资源的一种本质特性,主要指信息资源的利用不受人或人为干扰。不同用户可在同一时间或地点,或不同时间或地点共同利用同一种信息资源,而不需要任何的限制条件。信息资源共享的双方或多方均不会损失信息内容,相反还会产生新的信息。

(2) 知识性。知识性是一切社会信息资源所产生、传播的信息都具有的特性。人们在一定知识水平条件下,可以吸收社会信息,引起思维,继而利用它。

(3) 社会性。信息的存在、发展,可体现现代社会的本质和发展,它是人类社会活动必不可少的内容之一。信息资源的社会性既体现在信息的产生、存在、传播和使用上,还体现在信息资源的社会价值方面。

(4) 寄载性。信息必须借助于一定的符号存储与一定的载体中(包括人脑),才能被表现,没有载体,就没有信息。信息与载体,两者不能割裂开来。

(5) 增长性。与物质资源和能源资源的使用具有消耗性相比,信息资源的使用不仅不会减少信息资源的数量,而且在利用后,还会产生新的信息,使信息的数量不断增长。

(6) 动态性。信息资源是一种动态资源。信息资源产生于自然界和人类的社会实践活动之中,它随着时间的变化而变化,人类社会活动是一个永不停歇的过程,信息也总是处在不断产生、积累的过程中。

(7) 关联性。各种信息资源之间都存在各种复杂的关联,这是自然界的物质运动和生物活动过程中发生的各种关联以及人类社会活动中的各种关联关系的必然反映。

(8) 规模性。信息资源作为整体要有一定的量,分散、片面的信息不能较好地反映事物的情况。

## 二、信息资源类型

信息资源的类型,可根据多种依据来划分。

### 1. 以开发的程度为依据分类

以开发的程度为依据,可分为潜在的信息资源和现实的信息资源。

#### 1) 潜在的信息资源

潜在的信息资源是指人类在认识和思维创造的过程中,存储在大脑中的信息,只能为本人所利用,无法为他人直接利用,是一种有限再生的信息资源。

#### 2) 现实的信息资源

现实的信息资源是指潜在人脑中的信息通过特定的符号和载体表述后,可以在特定的条件下广泛地传递并连续反复地被人利用,是一种无限再生的信息资源。

### 2. 按表述的方式和载体为依据分类

按表述的方式和载体为依据,可分为口语信息资源、体语信息资源、实物信息资源和文献信息资源。

#### 1) 口语信息资源

口语信息资源是人类以口头方式表述但未被记录的信息资源,通常以讲演、授课、讨论等方式交流与利用;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) 体语信息资源

体语信息资源是人类在特定的文化背景下,以表情、手势、姿态等方式表述的信息资源,通常以表演、舞蹈等方式表现与交流;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

实物信息资源是人类通过创造性劳动以实物形式表述的信息资源,通常以样品、模型、雕塑等实物进行展示与交流;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

文献信息资源是人类用文字、数据、图像、声频、视频等方式记录在一定载体上的信息资源。特点是经过加工、整理,较为系统、准确、可靠,便于保存与利用,但也存在信息相对滞后,部分信息尚待证实的情况。只要这些载体不被损坏或消失,文献信息资源就可以跨越时空无限循环地为人类所利用,还可以按人类的需求整理成具有优化结构的文献信息资源体系。文献的一般分类有:

(1) 图书,是最古老、最重要的文献类型。国家标准《情报与文献工作词汇·传统文献》(GB13143—1991)对图书的解释是:一般不少于49页并构成一个书目单元的文献。按文种来分可分为:中文图书、日文图书、西文图书等;按作用范围可分为通俗图书、教科书、工具书等;按写作方式可分为专著、编著、翻译、编译等;按出版卷可分为单卷本、多卷本等;按刊行情况可分为单行本、丛书、抽印本等;按版次情况可分为初版、重版、修订本等。

(2) 连续出版物,是具有同一题名、定期或不定期以分册形式出版、有卷期或年月标识、计划无限期连续出版的文献,包括期刊、报纸、年度出版物及其他连续报告、会议记录、专著性丛刊等。连续出版物是与图书并列的最主要的文献类型,其特点是内容新颖、报道及时、出版连续、信息密集、形式一致等。据统计目前世界上连续出版物有130万种,限期连续出版物约50万种。

(3) 特种文献,是具有特定内容、特定用途、特定读者范围、特定出版发行方式的文献,包括学位论文、研究报告、专利、标准、产品样本、会议录、档案和政府出版物。国内外称这类文献为“灰色文献”,尽管这类文献并非很成熟,但它们内容新颖专深、实用性强、信息量大、参考性高、利用率大,是极为重要的信息资源。

(4) 网络信息资源,是以电子形式存贮于由成千上万台计算机组成的网络中的信息资源。它包括各类数据、电子文件、学术论文、图书、软件、商业活动等各种信息。因特网在20世纪70年代起源于美国,截至2000年底,已连接全球200多个国家和地区近5万多个网络、485万台主机、1.2亿个用户,并以每月20万个新用户的速度递增。

中国从1994年开始发展因特网,据《中国因特网发展状况统计报告》报道,截止2002年6月30日,中国上网计算机总数已超过1613万台,上网用户总数达到4580万,CN下注册的达126146个;WWW站点数(包括CN、COM、NET、ORG下的网站)约293213个。国外专家预计,中国将发展成为全球网上第一大用户,届时网上的中文信息也将越来越多,但在海量的信息中,重复交叉较多,垃圾信息也会很多。

根据不同的标准,可将网络信息资源划分成各种不同的类型:

按利用性质分,有开发性信息、注册式信息、交流式信息;按存取方式分,有邮件

型信息、电话型信息、揭示版型信息、广播型信息、图书馆型信息、数目型信息；按内容分，有商务信息、科技信息、社科信息、教育信息、娱乐信息等。

(5) 多媒体信息资源。将电信、电视、计算机三网相互融合，集图、文、声于一体的信息资源，包括网上广播电视、专题论坛、网上广告等。多媒体信息打破了图书、报刊、广播、电视单项媒体的界限，形成交互式媒体信息，可通过主题、文本、模版匹配，视频检索等方式对其进行检索。

### 第三节 信息检索基本原理

#### 一、信息检索概述

##### (一) 信息检索的含义与实质

信息检索又称为情报检索，萌芽于图书馆的参考咨询工作，20 世纪 50 年代才固定成专用术语。

信息资源检索是从任何信息集合中识别和获取所需信息的过程及其所采取的一系列方法和策略。从原理上看，它包括存储与检索两个方面，存储是检索的基础，检索是存储的反过程。而通常所说的信息检索是指狭义概念的信息检索，即从信息集合中找出所需信息的过程，也就是利用信息检索工具或数据库查找所需信息的过程。信息的检索过程则是信息存贮的逆过程。信息用户根据自己的需求对主题和概念进行认真分析后，将自己需检的信息集合。工作人员将标引后的信息条目录入，并将其按照一定的顺序排列起来，形成有序的信息集合——数据库，从而为信息检索奠定基础。要求转化为检索表达式，该检索表达式与系统标识的比较匹配过程就是检索的过程。

因此，信息检索的实质是将描述特定用户所需信息的提问特征，与信息存储的检索标识进行异同比较，从中找出与提问特征一致或基本一致的信息。

##### (二) 信息检索的重要意义与作用

信息检索的重要意义和作用主要体现在以下三方面：

##### 1. 能较全面地掌握有关的必要信息

信息检索可以有目的地、较系统地获得某一主题的必要信息。利用信息检索的方法充分了解国内外、前人和他人对拟探索或研究的问题已做过哪些工作？取得了什么成就？发展动向如何等。这样才能做到心中有数，防止重复研究，将有限的时间和精力用于创造性的研究中。

##### 2. 能提高信息利用的效率，节省时间与费用

一般公信度高的、较准确的信息才会被收集、组织或存储在检索工具或数据库中，有目的地查检检索工具所获得的必要信息比直接泛阅信息要快数十倍。



### 3. 能提高信息素质, 加速成才

信息素质是指具有获取信息的强烈意识, 掌握信息检索的技术和方法, 拥有信息鉴别和利用的能力。中国的高等教育明确要求大学生要具备信息素质。在当代社会, 人们需要终生学习, 不断更新知识, 才能适应社会发展的需求。美国工程教育协会曾估计, 学校教育只能赋予人们所需知识的 20%~25%, 而 75%~80% 的知识是走出学校后, 在研究实践和生产实践中根据需要不断再学习而获得的。因此, 掌握信息检索的方法与技能是形成合理知识和更新知识的重要手段, 是做到无师自通、不断进步的主要途径。

#### (三) 信息检索的类型

##### 1. 按存储和检索的内容划分

###### 1) 数据信息检索

数据信息检索是将经过选择、整理、鉴定的数值数据存入数据库中, 根据需要查询可回答某一问题的数据的检索类型。这些数据包括物理性能常数、统计数据、国民生产总值、外汇收支等。这类检索不仅查询数据, 还可以提供一定的推导、运算的能力。数据信息检索是一种确定性检索。

###### 2) 事实信息检索

事实信息检索是将存储于数据库中的关于某一事件发生的时间、地点、经过等情况查找出来的检索类型。它既包含数值数据库的检索、运算、推导, 也包括事实、概念等的检索、比较、逻辑判断。事实信息检索是一种确定性检索, 但事实信息检索过程中所得到的事实、概念、思想、知识等非数值性信息和一些数值性信息须进行分析、推理, 才能得到最终的答案, 因此要求检索系统必须具有一定的逻辑推理能力和自然语言理解功能。目前, 较为复杂的事实信息检索课题仍需人工才能完成。

数据检索和事实检索主要利用各种参考书来完成。

###### 3) 文献信息检索

文献信息检索通常指的是以二次信息(目录、索引、文摘)为工具的检索系统存储的信息检索类型, 这些二次信息是文献信息的外部特征与内容特征的描述集合体。文献检索是利用检索工具查出相关文献的过程。检索系统不直接解答用户提出的问题, 而是提供与之相关的文献名称及出处, 供用户筛选使用。检索结果是某本书、某篇文章、某份广告、某项专利或标准等一次文献。一些观点也认为可称为“数目检索”。

##### 2. 按组织方式分

###### 1) 全文检索

全文检索是将存储在数据库中的整本书、整篇文章中的任意内容信息查找出来的检索类型。可以根据需要获得全文中的有关章、节、段、句、词等的信息, 也可进行各种频率统计和内容分析。全文信息检索也是一种相关性检索, 它是在书目信息检索基础上更深层次的内容检索。随着计算机容量与运算速度的增大和提高, 全文检索正迅速由最

初的法律、文学领域扩大到更多的学科和专业。

### 2) 超文本检索

超文本检索是对每个节点中所存的信息以及信息链构成的网络中信息的检索类型。从组织结构上看,超文本的基本组成元素是节点和节点之间的逻辑连接链,每个节点中所存储的信息以及信息链被联系在一起,构成相互交叉的信息网络。强调中心节点之间的语义联结结构,靠系统提供的工具进行图示穿行和节点展示,提供浏览式查询,可进行跨库检索。

### 3) 超媒体检索

超媒体检索是对存储的文本、图像、声音等多种媒体信息的检索类型。它是多维存储结构,有向的链接,与超文本检索一样,可提供浏览式查询和跨库检索。

## 3. 按检索手段分

### 1) 手工检索

手工检索是人直接用手、眼、脑组织查找印刷型文献的检索类型。它是一种传统的信息检索,具有直观、灵活、无须各种设备和上机费用的优点。

### 2) 机器检索

机器检索又称计算机检索。它是一种现代的信息检索,是通过机器对已数字化的信息,按照设计好的程序进行查找和输出的过程。按机器检索的处理方式又分为脱机检索和联机检索;按存储方式又分为光盘检索和网络检索。机检可大大提高检索效率,拓宽检索领域。

## 4. 按是否使用检索工具分

### 1) 直接检索

直接检索是利用一次文献进行检索的方法,但所花时间多、精力大,检出文献少。这是以前比较常用的一种方法。

### 2) 间接检索

间接检索是指利用各种检索工具获得文献线索,再根据线索去查找原始文献线索的方法。

## 二、信息检索原理

信息检索的基本原理就是将检索提问标识与存储在检索工具中的标引标识进行比较,当检索标识与信息标引的标识相一致或部分一致时,即具有该标识的信息就从检索工具输出,输出的信息就是检索命中的信息。信息检索包括存储与检索两个部分。存储是对有关信息进行选择,并对信息特征进行著录标引和组织,建立信息数据库;检索则根据提问制定策略和表达式,利用信息数据库。从某个意义上说,信息存储是信息检索的逆过程,两者是不可分割的一个整体。无论手工检索还是计算机检索,其原理都是一样的,即对信息集合与需求集合的匹配与选择(图 1.1)。

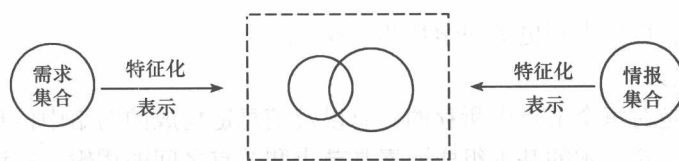


图 1.1 信息检索基本原理示意图

### (一) 信息检索效率

信息检索效率是研究信息检索原理的核心，是评价一个检索系统性能优劣的质量标准，它始终贯穿于信息存储和检索的全过程。衡量检索效率的指标有查全率、查准率、漏检率、误检率、响应时间等。目前，人们通常主要以查全率和查准率这两个指标来衡量。

#### 1. 查全率

利用检索系统进行某一课题检索时，检索出的相关信息量（ $w$ ）与该系统信息库中存储的相关信息量（ $x$ ）的比率再乘以 100%，称为查全率（ $R$ ），是对所需信息被检出程度的量度，用公式可表示为

$$R = w/x \times 100\% \quad (1.1)$$

式中： $R$ ——查全率；

$w$ ——检索出的相关信息量；

$x$ ——信息库中存储的相关信息量。

#### 2. 查准率

利用检索系统进行某一课题检索时，检出的相关信息量（ $w$ ）与检出信息总量（ $m$ ）的比率再乘以 100%，称为查准率（ $P$ ），即是衡量检索系统拒绝非相关信息的能力，用公式可表示为

$$P = w/m \times 100\% \quad (1.2)$$

式中： $P$ ——查准率；

$w$ ——检出的相关信息量；

$m$ ——检出信息总量。

从检索要求来说，希望查全率和查准率都同时达到 100%，即系统中存储的所有相关信息都被检索出（ $x=w=m$ ），这是最为理想的效果。但事实上很难达到全部检出和全部检准的要求，而只能达到某个百分比，总会出现一些漏检和误检。其漏检率（ $M$ ）和误检率（ $N$ ）也可用公式表示为

$$M = 1 - w/x \quad (1.3)$$

$$N = 1 - w/m \quad (1.4)$$

如果一个检索系统中与某一课题有关的信息共 250 条，实际检出 400 条，其中相关信息为 200 条，此次检索效率可计算为

查全率  $R = [200/250] \times 100\% = 80\%$ ，漏检率  $M = 1 - 80\% = 20\%$ ，

查准率  $P = [200/400] \times 100\% = 50\%$ ，误检率  $N = 1 - 50\% = 50\%$ 。