

信息检索与创新专利

(第二版)

刘二稳 许福运 编著
蔚莹 刘鹏 刘一农



科学出版社

信息 利

(第二版)

刘二稳 许福运
蔚莹 刘鹏 刘一农 编著

科学出版社

北京

内 容 简 介

本书以国内外信息检索系统的最新平台为素材,以增新删旧、易学易懂、使用方便为原则,揭示了国内外信息检索系统最新动态与使用方法。本书由信息资源检索概述、典型中文数据库检索、典型外文数据库检索、网络信息资源检索、专利及专利信息检索、创新专利步骤及案例、网络信息资源的综合利用,共7章内容组成。书中以国内外常用的网络数据库为代表,通过具体的检索实例,为学习者揭示国内外信息资源的检索方法与技能,并通过创新专利案例及训练、信息资源的综合利用来揭示创新专利实施的程序及信息资源的利用方法。

本书可作为普通高校本科生和研究生学习信息检索方法及技能的教材,也可作为教学、科研、工程技术和社会各界人士提高创新和检索能力的必备参考书。

图书在版编目(CIP)数据

信息检索与创新专利/刘二稳等编著. —2版. —北京:科学出版社, 2017.8

ISBN 978-7-03-053262-6

I. ①信… II. ①刘… III. ①信息检索 IV. ①G254.9

中国版本图书馆CIP数据核字(2017)第128543号

责任编辑:余江 张丽花/责任校对:郭瑞芝

责任印制:霍兵/封面设计:迷底书装

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>

三河市骏杰印刷有限公司印刷

科学出版社发行 各地新华书店经销

*

2013年8月第 一 版 开本:787×1092 1/16

2017年8月第 二 版 印张:12 1/4

2017年8月第八次印刷 字数:295 000

定价:32.00元

(如有印装质量问题,我社负责调换)

前 言

科学研究是站在巨人的肩膀上进行的创新，其整个过程离不开信息检索。信息检索是人们自主学习获取有效信息的有效途径，通过查询他人已有的成果，寻求答案激发灵感，做到知己知彼，从而在前人研究的基础上进行创新研究。本书从实际操作出发，讲授如何精准、快捷地获取自己想要的信息资源。掌握信息检索的技能，从容面对生活、学习、工作、科研中遇到的问题，从而帮助自己快速熟悉、掌握新知识、新领域，提高自己的学习能力和自身解决问题的能力，达到事半功倍的效果。

本书在第一版的基础上，以国内外信息检索系统的最新平台为素材，以内容全面、选材新颖、实用性和可操作性强为原则，揭示了国内外信息检索系统最新动态变化与使用方法。本书采用由浅入深、循序渐进的方式阐述信息检索的理论方法与实际操作过程，并通过具体的检索实例，让读者掌握其基本原理、检索技术、检索方法与流程，达到学以致用的目的。书中的网络信息检索系统均源自 2017 年的最新版本。由于版本的变化，内容上作了较大幅度的修改与补充，突出其实用性和可操作性。本书由信息资源检索概述、典型中文数据库检索、典型外文数据库检索、网络信息资源检索、专利及专利信息检索、创新专利步骤及案例、网络信息资源的综合利用共 7 章内容组成。各章节内容既自成体系又相互关联，可满足不同层次的读者需求。

本书第 1 章、第 2 章的 2.1 节~2.3 节、第 5 章、第 7 章由刘二稳编写；第 2 章的 2.4 节、2.5 节由蔚莹编写；第 3 章由刘鹏编写；第 4 章由刘一农编写；第 6 章由许福运编写。全书由刘二稳修改、统稿和定稿。

本书在编写过程中参阅和引用了许多同行的有关论著，出版过程中得到了科学出版社余江编辑的通力支持，在此一并表示衷心的感谢。

鉴于作者学术水平和选取的素材有限，加之时间仓促，书中难免存在疏漏之处，敬请读者批评指正。

作 者

2017 年 5 月

目 录

第 1 章 信息资源检索概述	1
1.1 信息资源	1
1.1.1 信息资源的含义及特性	1
1.1.2 信息资源的类型	1
1.2 信息检索的主要内容	7
1.2.1 信息检索的含义与原理	7
1.2.2 信息检索的类型	8
1.2.3 信息检索语言	9
1.2.4 信息检索途径	10
1.2.5 信息检索技术	11
1.2.6 信息检索程序	14
1.2.7 信息检索效果评价	16
1.3 图书馆文献的揭示与组织	17
1.3.1 馆藏文献的分类	17
1.3.2 馆藏图书的检索	19
第 2 章 典型中文数据库检索	22
2.1 中国知网	22
2.1.1 数据库简介	22
2.1.2 数据库检索	22
2.1.3 检索结果处理	26
2.1.4 检索实例	31
2.2 万方数据知识服务平台	32
2.2.1 数据库简介	32
2.2.2 数据库检索	33
2.2.3 检索结果处理	35
2.2.4 检索实例	36
2.3 维普期刊资源整合服务平台 7.0	38
2.3.1 数据库简介	38
2.3.2 数据库检索	39
2.3.3 检索结果处理	41
2.3.4 检索实例	43
2.4 汇雅电子图书	44
2.4.1 数据库简介	44

2.4.2	数据库检索	44
2.4.3	检索结果处理	46
2.4.4	检索实例	49
2.5	读秀学术搜索	50
2.5.1	数据库简介	50
2.5.2	数据库检索	51
2.5.3	检索结果处理	55
2.5.4	检索实例	58
第3章	典型外文数据库检索	60
3.1	SpringerLink 期刊全文数据库	60
3.1.1	数据库简介	60
3.1.2	数据库检索	60
3.1.3	检索结果处理	63
3.1.4	检索实例	63
3.2	EBSCO 期刊全文数据库	64
3.2.1	数据库简介	64
3.2.2	数据库检索	66
3.2.3	检索结果处理	69
3.2.4	检索实例	70
3.3	ScienceDirect 期刊全文数据库	70
3.3.1	数据库简介	70
3.3.2	数据库检索	71
3.3.3	检索结果处理	74
3.3.4	检索实例	76
3.4	Engineering Village 2 文摘数据库	76
3.4.1	数据库简介	76
3.4.2	数据库检索	76
3.4.3	检索结果处理	79
3.4.4	检索实例	81
3.5	Web of Science 数据库	82
3.5.1	数据库简介	82
3.5.2	数据库检索	83
3.5.3	检索结果处理	90
3.5.4	检索实例	92
第4章	网络信息资源检索	94
4.1	搜索引擎	94
4.1.1	搜索引擎的原理及类型	94
4.1.2	典型搜索引擎	95
4.2	免费网络学术资源搜索	106

4.2.1 免费图书检索	106
4.2.2 开放获取学术资源检索	109
4.3 网上标准文献资源检索	116
4.3.1 国内标准文献检索	116
4.3.2 国外标准文献检索	118
4.4 网上综合信息资源检索	119
4.4.1 教育信息资源检索	119
4.4.2 网上数据信息资源检索	121
第5章 专利及专利信息检索	124
5.1 专利概述	124
5.1.1 专利的概念	124
5.1.2 专利的类型	126
5.1.3 专利的申请	127
5.1.4 专利的分类	130
5.2 中国专利说明书种类及编号体系	131
5.2.1 中国专利说明书的种类	131
5.2.2 中国专利申请号、说明书及专利号编号体系	131
5.3 专利检索	132
5.3.1 专利检索途径及目的	132
5.3.2 中国专利信息检索	134
5.3.3 美国专利信息检索	138
5.3.4 欧洲专利信息检索	143
5.4 专利申请与专利授权样例	146
5.4.1 专利申请样例：多功能小太阳电暖器（发明专利）	146
5.4.2 专利授权样例：可办理银行卡的自动取款机（实用新型专利）	157
第6章 创新专利步骤及案例	160
6.1 创新步骤	160
6.1.1 提出创意	160
6.1.2 判断创新	161
6.1.3 完成创造	161
6.1.4 实现创效	162
6.2 创新专利案例	163
6.2.1 原始创新案例：光学视觉效应测试装置及测试方法	163
6.2.2 模仿创新案例：某公司进行风冷式冰箱的研发过程	168
6.3 创新专利训练案例	170
第7章 网络信息资源的综合利用	174
7.1 网络信息资源的收集、整理和分析	174
7.1.1 网络信息资源的收集方法	174
7.1.2 网络信息资源的整理方法	175

7.1.3 网络信息资源的分析方法	175
7.2 学位论文的选题与写作	176
7.2.1 学位论文的选题	176
7.2.2 学位论文的基本格式	178
7.3 写作与学术规范	180
7.3.1 写作规范	180
7.3.2 引用规范	180
7.3.3 学术规范	181
7.4 论文的收录、引用与发表	183
7.4.1 收录、引用	183
7.4.2 核心期刊与影响因子	184
7.4.3 核心期刊评价体系	184
7.4.4 论文投稿与发表	185
参考文献	188

第1章 信息资源检索概述

1.1 信息资源

1.1.1 信息资源的含义及特性

1. 信息资源的含义

信息资源是经人类主观或客观处理的,并且能够被传播或传输的文字、声音、图像、数据。信息资源有广义和狭义之分。广义的信息资源是指信息及其相关因素的集合。即除信息本身外,信息资源还包括与之紧密相关的信息设备、人员、系统、网络等,涉及信息的生产、分配、交换(流通)、消费等过程。狭义的信息资源是指信息本身或信息的集合,准确地说仅指信息内容。在国外有人将信息资源简单地分为四个组成部分:信息源、信息服务、信息产品和信息系统。

2. 信息资源的特性

人力资源、信息资源、物质资源是构成现代社会发展的三大支柱。信息作为一种资源,它除了具备物质资源的一般特征,如有用性、可选择性外,还具有以下特性。

共享性:这是信息资源区别于物质资源的根本属性之一。信息资源可同时为众多的使用者享用,也可被不同时期的使用者使用,还可为不同领域的人服务,满足他们不同的需要。它不像物质资源那样具有排他性。

时效性:即信息资源的价值对时间的灵敏度很高,一条及时的信息可能价值连城,一条过时的信息可能分文不值。

智能性:信息资源是人类按照一定的次序开发与组织起来的信息,是人类脑力劳动的产物,它传播的是人们的知识与智慧。

无限性与有限性并存:无限性是就信息资源的时间延续和储量而言的。信息资源产生于人类的社会实践活动,而人类的社会实践活动是一个永不停息的过程。因此信息资源也呈现出永不枯竭的特点。并且随着信息资源的不断开发和利用,信息资源也将日益丰富。有限性是指信息资源仅是信息的一部分,是经过人类选择的有用的那部分信息,是由人的智能的有限性决定的。

增值性:即对信息资源的投入和使用,不但可以使自然资源、人力资源、资本资源增值,而且可以使信息资源实现一次投入,多次开发利用,并不断产生新的信息使其自身增值。

1.1.2 信息资源的类型

信息资源按照不同的标准可划分成不同的类型。

1. 按文献载体的形式划分

文献载体是指记录知识的物质材料,随着科技的发展,人类记录知识信息的物质载体

和技术手段不断演进,相继出现了印刷型、缩微型、声像型和机读型等不同载体形式的文献。当前文献的载体主要是印刷型和机读型。

(1) 印刷型文献

印刷型文献是以纸张为存储介质,以印刷(包括铅印、胶印、静电复印等)为记录手段的文献形式,是一种传统的最常见的文献形式。其优点是便于携带、传播和阅读。缺点是体积大,存储密度低,长期保管困难。由于造纸材料(木材)的减少,因此印刷型文献的价格也越来越高。

(2) 机读型文献

机读型文献主要指将文字和图像转换成二进制数字代码,以计算机输入为记录手段,以磁带、磁盘或光盘为存储载体的文献,也称电子型文献。

电子型文献具有存储密度高,存取速度快,便于通信传递,易于复制共享等优点;其缺点是需借助计算机存储,设备价格较高,文件易丢失。

2. 按文献的出版形式划分

(1) 图书

图书(Book)又称为书籍,其内容比较成熟,按其出版形式可分为单卷书、多卷书、丛书等。图书内容成熟、系统,有完整定型的装帧形式的出版物。公开出版发行的图书一般标注有国际标准书号(ISBN)。

图书是对已有的科研成果与知识系统的全面概括和论述,并经过作者的认真核对、鉴别、筛选、提炼和融会贯通而成。从内容上,具有系统、全面、理论性强、成熟可靠、技术定型的特点;从时间上,由于编写时间、出版周期较长,所反映的文献信息的新颖性较差,但对要获取某一专题较全面、系统的知识,或对于不熟悉的问题要获得基本了解的读者,参阅图书是行之有效的方法。图书主要包括专著(Monograph)、丛书(Series of Monograph)、教科书(Textbook)、词典(Dictionary)、手册(Handbook)、百科全书(Encyclopedia)等。

图书的外表特征标识主要有书名(题名)、著者(责任者)、出版地、出版社、出版时间、版次、总页数、ISBN、价格等。

ISBN是International Standard of Book Number的缩写,即国际标准书号。ISBN是国际通用的图书或独立的出版物(定期出版的期刊除外)代码。出版社可以通过ISBN清晰地辨认所有书籍。ISBN具有唯一性。

2007年1月1日前的ISBN由10位数字组成,这10位数字分成4组,中间用“-”相连,每组数字都有不同的含义,即ISBN组号-出版社代码-书序码-计算机检验码。

第一组:地区号(国家、语言或区位代码),又叫组号,最短的只有一位数字,最长的达五位数字,大体上兼顾文种、国别和地区。0、1代表英语,使用这两个代码的地区有澳大利亚、加拿大、爱尔兰、新西兰、南非、英国、美国、津巴布韦等;2代表法语,法国、卢森堡、比利时、加拿大和瑞士的法语区使用该代码;3代表德语,德国、奥地利和瑞士德语区使用该代码;4是日本出版物的代码;5是俄罗斯出版物的代码;7是中国出版物使用的代码,香港的代号是962或988。

第二组:出版社代码。由国家或地区的ISBN中心设置并分给各个出版社。

第三组：书序码。该出版物代码，是出版者分配给每一个出版物的编号。

第四组：计算机校验码。校验码是 ISBN 的最后一位数字，它能够校验出 ISBN 是否正确。校验码只能是 1 位数，当为 10 时，记为罗马数字 X。

例如，《工程结构抗震》一书的 ISBN 为 7-144-05912-4，其中 7 为地区号或组号，代表中国；144 为出版社代码（人民交通出版社），取值范围 2~5 位数字，出版社的规模越大，出书越多，号码越短；05912 是书序号，由出版社给出，而且每个出版社的序号是定长的（数字 9 减去组号、出版社代码所占的位数，就是书序码的位数），最短的为一位，最长的为六位，出版社的规模越大，出书越多，序号越长，05912 表示人民交通出版社自使用 ISBN 编码以来出版的第 5912 种图书；4 代表计算机检验码，只有一位，为 0~9 或 X。

1982 年中国参加 ISBN 系统，并成立中国 ISBN 中心（设在国家新闻出版总署）。中国标准书号自 1987 年 1 月 1 日起实施。中国标准书号共分两部分：第一部分为 ISBN，是主体部分；第二部分为中国图书馆分类法基本大类类号和种次号。例如，ISBN 7-144-00316-X/TP·340，其中 TP 为中国图书馆图书分类法基本大类类号，代表计算机、自动化类；340 为种次号，代表该书为人民交通出版社自采用 ISBN 编号以来出版的第 340 种计算机、自动化类书。

为了扩大 ISBN 系统的容量，增加了另一个 EAN（国际物品编码协会）产品代码 979。2007 年 1 月 1 日以前，各国 ISBN 机构尚没有分配完的 10 位的 ISBN 可以加前缀 978，一旦现有的 10 位的 ISBN 用完了，新申请的 ISBN 全部以 979 开始。979 前缀目前主要用于两种情况：一是新建出版社首次申请 ISBN 和现有出版社在 2007 年 1 月 1 日以后申请 ISBN，二是如果以 978 为前缀的号码已用完，那么就用 979。

（2）期刊

期刊（Periodical 期刊，Journal 或 Magazine 杂志，Transaction 汇刊，Acta 学报，Serial），指有固定名称、定期或不定期出版、汇集了多位著者论文的连续出版物。与图书相比，期刊出版周期短，内容新颖，信息报道及时，是人们寻找研究上的新发现、新思想、新见解、新问题的首要信息源，也是科研人员特别是科技人员利用频率最高的一种文献信息源。

期刊的外表特征标识主要有期刊名、期刊出版者、出版地、出版年月日、期刊的卷号和期号、期刊的 ISSN（国际标准连续出版物号）、国内统一刊号（CN）、邮发代号、价格等。

期刊按照出版频率可分为周刊（Weekly）、半月刊（Semi-monthly）、月刊（Monthly）、双月刊（Bimonthly）、旬刊（3 Times a Month）、半年刊（Semi-annuals）、年刊（Annual Year Book）。

ISSN（International Standard Serial Number）是根据国际标准 ISO 3297 制定的连续出版物国际标准编码，其目的是使世界上每一种不同题名、不同版本的连续出版物都有一个国际性的唯一代码标志。该编号以 ISSN 为前缀，由 8 位数字组成。8 位数字分为前后两段各四位，中间用连接号相连，如 ISSN 1095-9203，前 7 位数字为顺序号，最后一位是校验码。

我国正式期刊的刊号由 ISSN 和 CN 两部分组成。国内统一刊号由 CN 和国内标准出版物编号组成，国内标准出版物编号由地区号、报刊登记号和中图分类号组成，如 CN31-1171/G2，地区号按《中华人民共和国行政区划代码 GB/T 2260—2013》取前两位，如北京为 11，天津为 12，上海为 31 等。国内内部刊物有准印号。

(3) 专利文献

专利文献 (Patent Document) 主要指专利申请人向自己国家或国外的专利局提出申请保护某项发明时所呈交的一份详细的技术说明书, 经专利局审查, 公开出版或授权后所形成的文献。专利文献具有技术含量高, 内容可靠, 实用性强, 报道快速, 能够最先反映新成果、新技术, 融技术信息、经济信息、法律信息为一体的特点。但专利文献的重复出版量大, 有时间、内容和地域方面的局限性。

专利文献的外表特征标识有申请号 (Application Number)、申请日期 (Application Date)、专利号 (Patent Number)、授权日期 (Granted Date)、优先号 (Priority Number)、优先国家 (Priority Country)、专利权人 (Assignee)、专利代理人、机构 (Attorney、Agent or Firm)、专利名称 (Title)、发明人 (Inventor)、权项 (Claims) 等。专利的文献内容特征包括专利分类号 (Patent Classification)、文摘 (Abstract)、检索领域 (Subject) 等。

(4) 会议文献

会议文献 (Conference Literature) 是指在国内外重要学术会议上发表的论文和报告。会议文献具有内容丰富、新颖、信息量大、专业性强、学术水平高、有创造性等特点。在主要的科技信息源中, 会议文献的重要性和利用率仅次于期刊。

会议文献的外表特征标识有论文名称、著者和著者工作单位、会议录名称、会议录出版情况、会议名称、会议时间、会议地点、会议届次等。会议文献的英文表达常用 Congress (会议)、Convention (大会)、Symposium (专题讨论会)、Seminar (学术研讨会)、Conference (学术讨论会)、Meeting (小型会议)、Proceedings (会议录) 等表示。

(5) 学位论文

学位论文 (英国 (Thesis), 美国 (Dissertation)), 是高等院校、科研机构的毕业生、研究生为了获取学位而撰写的学术性研究论文, 如博士学位论文、硕士学位论文、学士学位论文等, 其特点是具有学术性和独创性。大多数国家采用学士 (Bachelor)、硕士 (Master) 和博士 (Doctor) 三级学位制。通常学位论文主要指博士、硕士论文及优秀学士学位论文。

学位论文的外表特征标识有论文名称、作者、学位、授予学位的大学名称、时间、论文页码、导师或答辩委员会顾问的姓名等。

(6) 科技报告

科技报告 (Sci-Tech Report), 又称研究报告和技术报告, 是科技人员从事某一专题研究所取得的阶段性进展或最终性成果的记录与书面报告, 包括技术报告书 (Technical Report)、技术备忘录 (Memorandum book)、札记 (Note)、论文 (Paper)、通报 (Bulletin) 等。

科技报告的外表特征标识有报告名称、报告号、研究机构、完成时间等。

科技报告一般都有编号, 且单独成册。科技报告反映的是新兴科学和尖端科学的研究成果, 内容新颖, 专业性强, 能代表一个国家的研究水平, 各国都很重视。目前, 美、英、德、日等国每年产生的科技报告达 20 万件左右, 其中美国占 80%。美国政府的 AD、PB、NASA、DOE 报告是国际上最著名的四大科技报告。

PB (Publishing Board) 报告: 产生于第二次世界大战结束之后, 当时美国政府为了整理和利用从战败国获得的数以千吨计的秘密科技资料, 于 1945 年 6 月成立了一个专门的出

版局,即美国商务部出版局(Publication Board),负责收集、整理、报道利用这些资料,报告的内容侧重于各种民用科学技术、生物医学。

AD(ASTIA Document)报告:产生较早,1951年开始出版。原来为美国武装部队技术情报局(Armed Services Technical Information Agency, ASTIA)出版的文献,即ASTIA Document报告。现由美国国防技术情报中心(Defense Technical Information Center, DTIC)负责收集整理和出版。报告内容以国防部各个合同户的研究报告为主,其报告号冠以AD, A原先为Armed、D为Document之意,现在则可理解为入藏报告(Accession Document),主要收录军事科技方面的文献资料。

NASA报告:是美国航空航天局(National Aeronautics and Space Administration)的简称。内容除航空航天技术以外,还涉及许多相关学科,在一定程度上成为综合型科技报告。

DOE报告:是美国能源部(US Department of Energy)出版的报告,收录能源部所属实验室、能源技术中心和情报中心以及合同单位发表的科技报告,内容涉及核能与其他能源,包括矿物燃料、太阳能以及节能、环境和安全等内容。

科技报告具有保密的特点,因而不易获取。我国国家图书馆、国防科技信息研究所和上海图书馆的科技报告相对比较完整。

(7) 标准文献

标准文献(Standard Literature)指对工农业新产品和工程建设的质量、规格、参数及检验方法所作的技术规定,是人们在设计、生产和检验过程中共同遵守的技术依据。它是一种规章性的技术文件,具有一定的法律约束力。按批准机构级别和适用等级可分为国际标准、国家标准、部颁标准(行业标准)和企业标准四个等级。国际标准指国际间通用的标准,如ISO(国际标准)、IEC(国际电工标准)、IEEE(美国电气和电子工程师协会)等。我国目前将标准分为国家标准、行业标准、地方标准和企业标准四级。我国标准有效期一般为5年。

标准文献都有标准号,它通常由国别(组织)代码+顺序号+年代组成。我国的国家标准分为强制性国标(GB)和推荐性国标(GB/T);行业标准代码以主管部门名称的汉语拼音声母表示,如JT表示交通行业标准。企业标准编号:Q/省、市简称+企业名代码+年份。

标准一般有如下特点:①每个国家对于标准的制定和审批程序都有专门的规定,并有固定的代号,标准格式整齐划一;②在一定条件下具有某种法律效力,有一定的约束力;③时效性强,不断地进行修订、补充、替代或废止;④一个标准一般只解决一个问题,文字准确简练;⑤不同种类和级别的标准在不同范围内贯彻执行。

标准文献的外表特征标识有标准级别、标准名称、标准号、标准提出单位、审批单位、批准时间。

(8) 政府出版物

政府出版物(Government Publication)是指各国政府及所属机构发表的文件,分为行政性和科技性两大类。行政性文件包括政府报告、会议记录、法令、条约、决议、规章制度、调查统计资料等;科技性文件包括科研报告、科普资料、科技政策、技术法规等。政府出版物的特点是具有正式性和权威性。根据其性质分为公开资料、内部资料、机密资料三种。

(9) 产品资料

产品资料(Product Literature)是国内外生产厂商或经销商为了推销产品而印发的企业出版物,用来介绍产品的情况,以产品目录、产品样本、产品说明书、企业介绍等多种形式对外宣传。产品资料反映的技术可靠成熟、图文并茂、直观形象、数据翔实,是宝贵的科技信息源、商贸信息源和竞争情报源,免费赠送的较多,具有简明可靠、形象直观、时间性强、形式多样、出版迅速等特点。

(10) 技术档案

技术档案(Technology Archives, Technical Records)指生产建设、科技部门和企事业单位针对具体的工程或项目形成的技术文件、设计图样、图表、照片、原始记录的原本及复印件,包括任务书、协议书、技术经济指标和审批文件、研究计划、研究方案、试验记录等。它是生产领域、科学实践中用以积累经验、吸取教训和提高质量的重要文献。科技档案具有保密性,常常限定使用范围。

3. 按加工的深度划分

从文献内容出发,根据对文献信息加工处理的深度不同,文献可分为零次文献、一次文献、二次文献和三次文献,不同加工深度的文献特点鲜明,不同的检索需求需要检索不同加工深度的文献。

(1) 零次文献

零次文献(Zeroth Literature)指未经正式出版或非正式渠道交流的文献,未公开于社会,只为个人或某一团体所用,如文章底稿、私人笔记、书信、工程图纸、考察记录、实验记录、调查稿、原始统计数字,以及各种口头交流的知识、经验、意见、论点等。其主要特点是传递速度快、针对性强,但不成熟、传播面窄、不公开交流、不易获取和管理。

(2) 一次文献

一次文献(Primary Literature)指以作者的研究成果为基础而撰写的原始文献,包括学术论文、专著、科技报告、专利说明书、标准文献、档案等公开发表的类型。其特点是形态多样,因而同一主题内容的形式很分散;内容原创新颖、详尽、系统,参考和使用价值大,是检索和利用的主要对象。

(3) 二次文献

二次文献(Secondary Literature)又叫检索性文献,习惯上又称检索工具,二次文献是通过对一次文献进行外部特征和内容特征的分析、提取、整理而形成的新的文献形式。一般包括目录(Bibliography)、题录(Title)、索引(Index)、文摘(Abstract)等。二次文献不仅能报道一次文献,还为查找一次文献提供线索,又是检索一次文献的主要工具。其汇集性、工具性特征突出,提供的文献线索集中、系统、有序。

①目录。目录又称书目(Bibliography),是对出版物的外表特征加以揭示与报道并按照一定的顺序编排而成的,主要提供书刊的出版信息、收藏信息,以备订购、宣传和查阅。目录是各图书情报单位报道文献的主要方式,书目分为国家书目、专题书目、推荐书目、馆藏目录和联合目录等。其中,馆藏目录反映单个图书馆的书刊收藏情况,联合目录反映的是多个图书馆的书刊收藏情况。

②题录与文摘。题录是描述单篇文献外部特征的条目集合。题录包含文献标题、作者、

作者工作单位、发表时间、文献来源（期刊、会议、专利等）。文摘也称摘要，以提供内容梗概为目的，言简意赅，准确反映原文的研究目的、观点或论点、方法、数据、结论等主要内容。

③索引。索引是对一组信息集合的系统化的指引。索引一般只起指引特定信息内容及存储地址的作用。一种索引就能提供一种检索途径，形成一个检索点或检索入口。索引按揭示信息的外部特征分为著者索引（Author Index）、号码索引（专利号索引（Patent Number Index）、报告号索引（Report Number Index））等；按揭示信息的内容特征可分为主题索引（Subject Index）、分类索引（Classification Index）、关键词索引（Keyword Index）等。用户可以根据检索课题的已知信息利用数据库提供的相应索引进行检索。索引按揭示的对象可分为期刊索引（Periodicals Index）、会议录索引（Proceedings Index）、引文索引（Citation Index）、报纸索引（Newspaper Index）等。

（4）三次文献

三次文献（Tertiary Literature）又叫参考性文献，是通过二次文献提供的线索，对一定范围内的一次文献的内容进行分析、研究，综合后编成的文献。三次文献可分为综述研究类和参考工具类两种类型，前者如动态综述、学科总结、专题述评、进展报告等；后者如年鉴、手册、大全、词典、百科全书、教科书等。

1.2 信息检索的主要内容

1.2.1 信息检索的含义与原理

1. 信息检索的含义

检索即查找、索取，是根据特定的需求，运用检索工具，按照一定的方法，从大量信息中查出所需信息的工作过程。

信息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。从信息检索的含义可以看出，它包括信息存储与检索两个方面，信息存储形成信息集合，它是检索的基础，检索是从信息集合中查找特定的信息，它是存储的逆过程。

信息存储过程——对大量无序的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，即形成检索系统（数据库）的过程。

信息检索过程——根据用户的信息需求，利用检索系统（数据库），选择检索系统提供的方法与途径检索有关信息，即检索系统的使用过程。

2. 信息检索原理

各种检索系统的检索原理基本相同，简单地讲就是对信息集合与需求集合的匹配与选择，如图 1-2-1 所示。

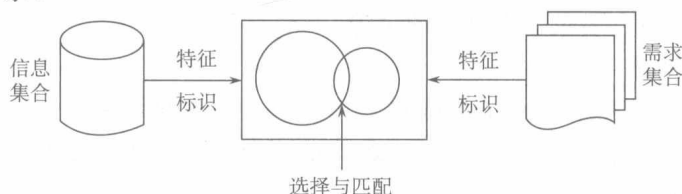


图 1-2-1 信息检索原理

因此,信息检索的本质是将用户所需信息的检索提问标识与检索系统中的信息标引的标识进行对比,从中找出与提问特征标识一致或基本一致的信息。提问特征标识是对信息的需求进行分析,从中选择能代表信息需求的主题词、分类号或其他符号。

要保证信息检索标识与信息标引标识能够匹配对比,存储与检索所依据的规则必须一致,即标引者与用户必须遵守相同的标引规则。这样无论什么样的标引者,只要对一篇文献的标引结果一致,不论谁检索,都能查到这篇文献。例如,标引者用分类号标引文献,检索者也必须用分类号检索,否则二者无法匹配对比。

1.2.2 信息检索的类型

检索是对未知知识最有效的获取方法,也是人们根据需要,利用人类有史以来积累的知识的唯一有效方法。它的最大优势在于可以用最少的时间以最快的速度获得尽可能多的信息与知识。文献信息检索根据划分标准不同分为不同的检索类型。

1. 按信息存储与检索方式划分

(1) 手工检索

手工检索简称手检,是指人们通过手工的方式检索信息。其使用的检索工具主要是书本型、卡片式的信息系统,即目录、索引、文摘和各类工具书。检索过程是由人工以手工的方式完成的。

(2) 计算机检索

计算机检索简称机检,是指人们利用数据库、计算机软件技术、计算机网络及通信系统进行的信息检索,其检索过程是在人机的协同作用下完成的。

(3) 综合检索

在文献信息检索的过程中,既使用手工检索方式,又使用计算机检索的方式,也就是同时使用两种检索方式。

2. 按检索内容(或检索目标)划分

(1) 数据信息检索

数据信息检索是一种确定性检索,是以数值或图表形式表示的数据为检索对象的信息检索,又称数值检索。检索系统中存储的是大量的数据,这些数据既包括物质的各种参数、电话号码、银行账号、观测数据、统计数据等数字数据,也包括图表、图谱、市场行情、化学分子式、物质的各种特性等非数字数据。

(2) 事实信息检索

事实信息检索是以特定客观事实为检索对象,借助提供事实检索的检索工具与数据库进行检索。其检索结果为基本事实,如某个字、词、诗词文句、地名、人物、某一机构、某一事件的查找等。一般来说,事实检索多利用词语性和资料性工具书,包括字典、词典、百科全书、年鉴、手册、名录、表谱、图录等。

(3) 文献信息检索

文献信息检索是指利用检索工具或检索系统查找文献的过程,包括文献线索检索和文献全文检索。文献线索检索是指利用检索工具或检索系统查找文献的出处,检索结果是文献线索。它包括书名或论文题目、著者、出版者、出版地、出版时间等文献外部特

征。用于检索文献线索的检索工具有书目、索引、文摘及书目型数据库和索引、题录型数据库。

3. 按信息的组织方式划分

(1) 全文检索

全文检索是以文献所含的全部信息作为检索内容，即检索系统存储的是整篇文章或整部图书的全部内容。检索时可以查到原文及有关的句、段、节、章等文字，并可进行各种频率统计和内容分析。全文检索主要是用自然语言表达检索课题，较适用于某些参考价值大的经典性文章，全文检索是计算机信息检索的发展方向之一。

(2) 超文本检索

超文本结构类似于人类的联想记忆结构，它采用一种非线性的网状结构组织块状信息，没有固定的顺序，也不要求读者必须按照某个顺序来阅读。采用这种网状结构，各信息块很容易按照信息的原始结构或人们的“联想”关系加以组织。

(3) 超媒体检索

把多媒体信息引入超文本里就产生了多媒体超文本，即超媒体。它是对超文本检索的补充，其存储对象超出了文本范畴，融入了静态、动态图像及声音等多媒体信息。信息存储结构从单维发展到多维，存储空间范围不断扩大。

1.2.3 信息检索语言

当用户有信息需求或遇到疑难问题时，就需要检索相关信息，检索前，用户首先要分析自己的信息需求，形成检索提问标识，而且这个检索提问标识必须与检索系统的标引标识一致，相关的文献才能被检索出来，这种保证存储与检索一致的语言就是检索语言，检索语言是信息存储与检索共同遵循的规则。检索语言是描述文献特征、用于标引和检索的人工语言，是沟通标引与检索的桥梁。检索语言分类如图 1-2-2 所示。

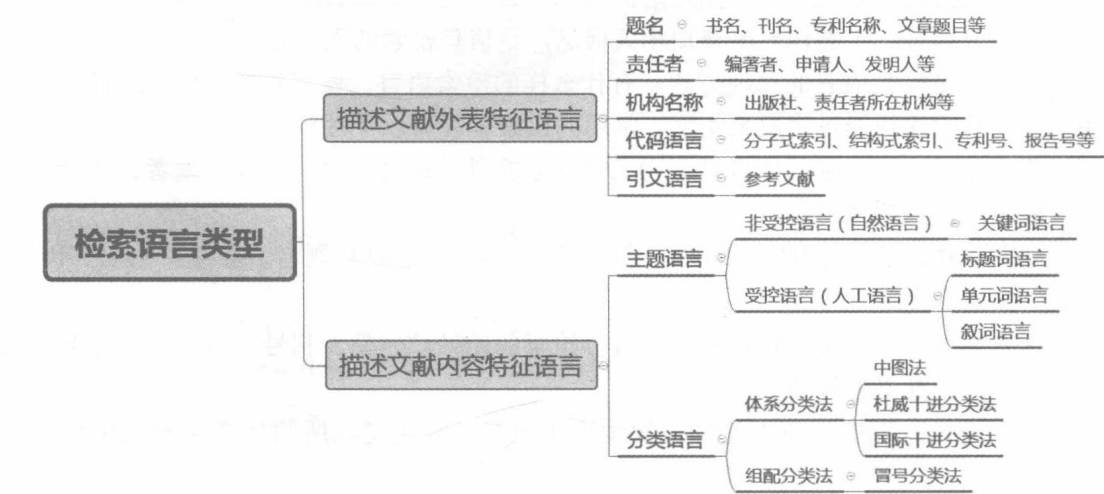


图 1-2-2 检索语言分类