

数字文献检索

舒炎祥 方胜华 编 著



科学出版社

www.sciencep.com

数 字 文 献 检 索

舒炎祥 方胜华 编著

科 学 出 版 社

北 京

内 容 简 介

本书系统地介绍了数字文献检索的基础知识,包括检索语言、计算机检索技术、检索策略、检索效果的评价等。还对国内外的期刊论文全文数据库、文摘数据库、综合检索平台、电子图书、电子报纸、专利、会议论文、学位论文、科技报告、标准文献数据库、参考数据库等数字文献资源的基本概况和检索方法进行了介绍。此外,还对互联网上一些著名搜索引擎的用法、技巧作了介绍。本书基本反映了国内外主要的数字文献资源,内容编排结构合理,数据库内容新颖实用,检索途径介绍得全面系统。

本书可作为高等院校本科生、研究生的信息检索教材,也可作为科研人员、图情人员的培训教材及参考用书。

图书在版编目(CIP)数据

数字文献检索 / 舒炎祥, 方胜华编著. —北京: 科学出版社, 2010

ISBN 978-7-03-026304-9

I. 数… II. ①舒…②方… III. 数字技术—应用—情报检索 IV. G252.7-39

中国版本图书馆 CIP 数据核字 (2009) 第 241766 号

责任编辑: 陈晓萍 / 责任校对: 王万红

责任印制: 吕春珉 / 封面设计: 东方人华

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>

新蕾印刷厂印刷

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: 11 1/2

印数: 1—4 500 字数: 272 000

定价: 20.00 元

(如有印装质量问题, 我社负责调换〈新蕾〉)

销售部电话 010-62134988 编辑部电话 010-62138978-8003

版权所有, 侵权必究

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

前 言

高校“文献检索”课程的开设,对于培养大学生的信息素养,提高大学生的信息获取能力和创新能力具有不可替代的积极作用。

但随着计算机技术和互联网发展,数字文献的数量和种类正变得无限丰富,不断有新的数据库产品出现。读者获取信息的对象从纸质文献转身数字文献,获取的方法、途径和手段等也发生了巨大的变化。在网络环境下查找数字文献需要知道两个问题:一是有什么?即有哪些数字文献可供我们进行检索,我们查找具体的信息时,又应该到哪些具体类型的数字文献去查找呢;二是怎么找?即对这些数字文献怎么检索,如何方便、快捷、准确和全面地获取所需的数字文献信息呢。本书正是为了解决这两个问题,同时适应高校教学改革和人才培养目标的需要,帮助高校学生学习掌握网络环境下数字文献的检索而编写的一本通用性和针对性相结合、编排结构新颖的文献检索教材。

全书共九章,分为四个部分。第一部分(第1章)主要介绍数字文献的基本概况、数字文献的检索技术以及检索效果的评价等,让读者了解数字文献检索的基础知识和基本技术。第二部分(第2章至第7章)介绍了中国期刊论文全文数据库、国外期刊论文全文数据库、文摘/题录数据库与综合检索平台、电子图书、电子报纸、特种文献、参考数据库等数字文献数据库的基本概况以及检索方法。第三部分(第8章)介绍了网络搜索引擎的基础知识,以及两个专业搜索引擎的用法。第四部分(第9章)介绍了科研论文、文献综述和文摘的写作基础知识。全书各章节的结构和部分内容参考了章云兰、万跃华、舒炎祥编著的《数字资源检索教程》一书,在此表示衷心的感谢。

本书由舒炎祥、方胜华编著,编写分工如下:陈森编写第1章,舒炎祥编写第2、5章,叶升阳编写第3章,宋少平编写第4章,方胜华编写第6章,刘金编写第7章,邹丽红编写第8章,王静芳编写第9章。全书由舒炎祥进行统稿和定稿。

由于编写者的水平有限,加之时间仓促,书中难免会有疏漏、错误之处,敬请同行、师生和读者批评指正。

舒炎祥

2009年12月

目 录

前言

第 1 章 数字文献检索基础	1
1.1 数字文献及其检索概述	1
1.1.1 文献及数字文献概述	1
1.1.2 数字文献的检索语言	7
1.2 数字文献的检索技术	11
1.2.1 布尔逻辑检索技术	12
1.2.2 截词检索技术	13
1.2.3 位置检索技术	14
1.2.4 字段检索技术	15
1.3 数字文献的检索策略	16
1.3.1 信息需求分析	16
1.3.2 数据库的选择	17
1.3.3 检索词的确定	18
1.3.4 检索表达式的构造	18
1.4 检索效果评价	19
1.4.1 概述	19
1.4.2 查全率	19
1.4.3 查准率	20
本章主要参考文献	21
第 2 章 中国期刊全文数据库	22
2.1 中国期刊全文数据库	22
2.1.1 概述	22
2.1.2 开始使用数据库	24
2.1.3 数据库检索	25
2.1.4 检索结果处理	31
2.2 中文科技期刊数据库	32
2.2.1 概述	32
2.2.2 开始使用数据库	34
2.2.3 数据库检索	34
2.2.4 检索结果显示及全文下载	39
2.2.5 我的数据库	41
2.3 万方数据资源系统	41
2.3.1 概述	41

2.3.2 数据库检索	43
2.3.3 个性化服务	48
2.4 人大复印报刊资料数据库	48
2.4.1 概述	48
2.4.2 数据库检索	49
本章主要参考文献	50
第3章 国外期刊论文全文数据库	51
3.1 Elsevier Science 出版社的电子期刊	51
3.1.1 概况	51
3.1.2 数据库检索	52
3.2 Springer 出版社的电子期刊	56
3.2.1 概况	56
3.2.2 数据库检索	57
3.3 World Scientific Publishing 的电子期刊	59
3.3.1 概况	59
3.3.2 数据库检索	60
3.4 EBSCOhost 系统全文数据库	64
3.4.1 概况	64
3.4.2 数据库检索	65
3.5 LexisNexis Academic 学术大全数据库	73
3.5.1 概况	73
3.5.2 数据库检索	74
本章主要参考文献	78
第4章 文摘、题录数据库与综合检索平台	79
4.1 科学引文索引 (SCI)	79
4.2 工程索引 (Ei)	81
4.3 化学文摘 (CA)	82
4.4 SCOPUS	84
4.5 DIALOG 国际联机检索系统	85
4.6 全国报刊索引数据库	86
4.6.1 概述	86
4.6.2 数据库检索	87
4.7 中国高校人文社会科学文献中心 (CASHL)	90
4.8 综合检索平台	91
4.8.1 国家科技图书文献中心检索平台 (NSTL)	91
4.8.2 中文社会科学引文索引 (CSSCI)	93
本章主要参考文献	94
第5章 电子图书和电子报纸	95
5.1 电子图书	95

5.1.1 电子图书概述	95
5.1.2 中文电子图书服务系统	96
5.1.3 外文电子图书	102
5.2 电子报纸	104
5.2.1 电子报纸概述	104
5.2.2 中文电子报纸数据库	105
5.2.3 外文电子报纸数据库	106
本章主要参考文献	107
第6章 特种文献数据库	108
6.1 专利文献及其检索	108
6.1.1 专利与专利文献概述	108
6.1.2 中国专利信息的检索	109
6.1.3 国外专利信息的检索	111
6.2 会议文献及其检索	118
6.2.1 会议文献概述	118
6.2.2 国内会议文献的检索	119
6.2.3 国外会议文献的检索	119
6.3 学位论文及其检索	120
6.3.1 学位论文概述	120
6.3.2 国内学位论文的检索	120
6.3.3 国外学位论文的检索	121
6.4 科技报告及其检索	121
6.4.1 科技报告概述	121
6.4.2 万方数据的《中国科技成果数据库》	122
6.4.3 美国四大科技报告的检索	122
6.5 标准文献及其检索	123
6.5.1 标准文献概述	123
6.5.2 中国标准的检索	123
6.5.3 国外标准的检索	124
本章主要参考文献	127
第7章 参考数据库	129
7.1 中文参考数据库	129
7.1.1 中国资讯行: 高校财经数据库	129
7.1.2 中经网: 教育版	131
7.1.3 国研网: 教育版	133
7.1.4 新华社多媒体数据库: 高等教育版	134
7.2 英文参考数据库	136
7.2.1 贝尔斯坦/盖墨林化学数据库	136
7.2.2 ISI 化学数据库	138

7.3 其他类型参考数据库	140
7.3.1 多媒体视频系统	140
7.3.2 英语学习系统	143
7.3.3 考试系统	145
本章主要参考文献	147
第8章 网络资源的搜索	148
8.1 综合搜索引擎	148
8.1.1 搜索引擎概述	148
8.1.2 百度搜索引擎	151
8.1.3 Google 搜索引擎	154
8.2 专业搜索引擎	156
8.2.1 Scirus 学术搜索引擎	156
8.2.2 同方知网的知识搜索	158
8.3 搜索引擎使用技巧	160
8.3.1 搜索引擎搜索技巧	160
8.3.2 网络学术资源的其他收集办法	161
8.3.3 著名搜索引擎网址	162
本章主要参考文献	165
第9章 科研论文的写作	166
9.1 科研论文的写作	166
9.2 文献综述的写作	172
9.3 文摘的写作	174
本章主要参考文献	175

第1章 数字文献检索基础

1.1 数字文献及其检索概述

1.1.1 文献及数字文献概述

1. 文献的概念

世界上最早的文献产生于距今大约 10000~40000 年前的旧石器时代。雕刻在洞壁上的石刻图像、古代西亚的楔形文字文献、古埃及的纸草文献、古希腊克里特线性文字文献和中国殷商时期的甲骨文献等，都是世界上最古老的文献。

在我国，“文献”一词始见于《论语·八佾》：“夏礼吾能言之，杞不足徵也；殷礼吾能言之，宋不足徵也；文献不足故也。足，则吾能徵之矣。”宋代著名学者朱熹在《四书章句集注》中注释为：“文，典籍也；献，贤也。”

目前，我国颁布的《中华人民共和国国家标准——文献著录总则》（GB3792.1—1983）中文献的定义为“记录有知识的一切载体”。具体地说，文献是用文字、图表、符号、图像、音频和视频等记录知识的物质形态。

文献由知识、载体和记录方式 3 个要素构成。知识是内容，载体是形态，记录方式是构成文献的手段，三者不可分割。

2. 文献的类型

根据划分标准的不同，文献有多种分类方式。

（1）按文献出版形式和内容划分

1) 图书。图书是指对某一领域的知识进行系统阐述或对已有研究成果、技术以及经验等进行归纳、概括，是一种比较系统完整而又成熟定型的出版物。国际上通常将 48 页以上、不定期出版的印刷品称为图书。

图书一般具有完整定型的装帧形式，首尾衔接，结构严谨，以单行本、多卷书或丛书等形式出版发行。图书的内容系统、论述全面；反映的知识可靠、观点成熟。但传统的印刷图书出版周期长、更新速度慢，而电子图书的出现逐渐弥补了这一缺陷。

2) 期刊、报纸。期刊又叫做杂志，指有固定的名称、版式、开本和篇幅，汇集若干作者分别撰写的多篇文章、资料或线索，由常设编辑机构的工作人员依照一定的出版周期，使用连续的卷期号或年月顺序号作为时序的标识，计划无限期连续出版的出版物。

期刊具有内容广泛、知识新颖、出版周期短、出版量大、传递信息快以及流通范围广等特点，读者可以很方便地获得自己所需要的期刊，及时了解国内外各学科领域最新研究成果和动态，是科技人员进行信息交流的正式、公开而有序的工具。据统计，科技人员所获取信息的 65% 以上来源于期刊，它是十分重要和主要的信息源和检索对象。随着互联网和计算机技术的发展，许多印刷型期刊也同时有电子版发行，还出现了纯电

子版期刊, 同时还有各种全文和检索性数据库, 使期刊的检索和利用更加方便。

报纸和期刊一样, 有统一的名称, 定期连续出版, 每期汇集许多文章、报道、资料 and 消息。报纸比期刊时间性更强, 出版周期更短, 有日报、周报和旬报等不同出版周期之分, 是信息报道最及时的一种文献; 内容更加广泛, 它以最快的速度报道世界各地发生的时事新闻、评论和各学科领域出现的最新成果。报纸是重要的情报源和社会舆论工具, 对社会经济生活有着广泛的影响。但是, 报纸报道信息的系统性和完整性不如图书、期刊等其他文献类型。

3) 科技报告。科技报告是关于某项科学研究的系统总结, 或是研究过程中阶段进展情况的实际记录, 又称技术报告。它通常以正式报告、进展报告、技术札记与备忘录等形式出版。

科技报告的内容专深具体, 常附有大量的数据、图表与原始实验记录等材料, 许多科技报告与政府的研究活动、国防及尖端科技领域有关, 有一定保密性。科技报告所报道的研究成果一般经过有关部门的审查和鉴定, 所反映的技术内容较为成熟, 数据较为详尽、可靠, 报道及时, 具有很高的情报价值, 是一种重要的信息源。科技报告发行范围被严格控制, 获取比较困难。

4) 政府出版物。政府出版物指各国政府及其所属机构颁布的文件。它的内容广泛, 可分为行政性文件和科技文件两大类, 其中行政性文件包括会议记录、各种法令、决议、方针政策、外交文件、统计数据和规章制度等; 科技文件包括各部门的研究报告、政府颁布的科技政策、条例等。政府出版物是了解各国政治、法律、军事、经济、科技、文化和教育等不可缺少的权威性信息来源和重要参考资料。

5) 会议文献。会议文献指学术会议文献, 包括会前、会中和会后文献。会前文献指会议日程表、会议论文预印本和论文摘要等; 会中文献指开幕词、讨论记录、会议决议和闭幕词等; 会后文献是指会议录、会议论文集、会议论文汇编、期刊特辑及有关会议的声像资料等。其中会后文献内容比较系统完整, 是会议文献的最主要部分, 但会后文献没有固定的出版形式。

科学的发展离不开交流、讨论及争鸣, 而召开学术会议, 可以使相关学者会聚一堂当面探讨, 是最佳的交流方式。各学科的前沿动向、研究热点、发展趋势及最新的学术思想、科研成果和实验方法都可在会议上充分展示, 不同的学派、不同的观点以及不同的研究风格会在会议上激烈碰撞。据统计, 会议文献对特定领域的重大事件和发现的首次报道率最高, 但其内容与期刊相比可能不太成熟, 是了解新动向、新发现的重要信息源。

6) 学位论文。学位论文是高等学校或研究机构的学生为了取得学位而撰写的学术论文。学位论文的水平差异较大, 具有一定学术质量的硕士或博士学位论文, 是学生在导师的指导下, 经过长期调查研究, 收集大量文献资料, 就某一课题作出全面系统的论述, 提出独创性的见解, 可能具有较高的学术水平和参考价值, 是重要的参考文献资料。

学位论文是非卖品, 一般也不公开发表, 通常只在学位授予单位的图书馆和按国家规定接受呈缴本的图书馆保存有副本, 故学位论文的收集与利用不如其他类型的文献方便。随着信息化程度的提高, 许多机构同时保存学位论文的文本型和电子型样本, 还把本机构的学位论文作为特色资源, 建成学位论文检索系统和全文数据库供本单位读者利用, 也有专门的数据库厂商进行学位论文的收录和反映。

7) 专利文献。专利是指建立专利制度的国家通过其政府机构(专利局)以法律形式保护发明人在一定时期内享有的技术专有权利。专利文献是专利制度的产物。广义的专利文献是指所有与专利制度有关的文件,包括专利申请书、专利说明书、专利公报、专利分类表、专利检索工具、专利证书以及与专利相关的其他文献;狭义的专利文献一般指各国专利局出版的各种发明说明书或专利说明书及其派生出来的二次文献,它是专利文献的主体。

专利文献具有编写格式统一、出版快、内容新颖、技术性强、实用性强并具有法律效力等特点,是制订科研规划、规划产品组合战略、确定工艺路线、实施技术改造的一个主要技术信息源。专利文献的电子版发展较快,特别是专利检索,我国的专利和许多国家的专利都有相应的网站可以免费检索,大多数专利文献同时有印刷版和电子版。

8) 标准文献。标准是对重复性事物和概念所做的统一规定,它以科学、技术和实践经验的综合成果为基础,经有关方面协商一致,由主管机构批准,以特定形式发布,作为共同遵守的准则和依据。换言之,标准是技术标准、技术规范和技术法规的总称。标准文献是指与标准化活动有关的一切文献,是反映标准的技术文献。

标准文献具有针对性、时效性。它所反映的水平只能是当时技术所能达到的水平,随着经济发展和科学技术水平的提高,标准要不断进行修订、补充、替代或废止。一个标准一般只解决一个问题,不同种类和级别的标准只能在其适用的范围内贯彻执行。标准文献反映了这个国家的经济政策、技术政策、生产水平、标准化程度和资源开发程度,对产品更新换代、改进工艺水平、提高产品质量和加强市场竞争力可以起到很好的指导作用。

9) 产品样本。产品样本是厂商为向客户宣传和推销其产品而印发的介绍产品情况的文献。其内容主要是对产品的规格、性能、特点、构造、用途和使用方法等进行介绍和说明,包括产品样本、产品说明书、产品目录与厂商介绍等。

产品样本的特点在于技术成熟、数据可靠,有较多的外观照片和结构图,直观性强,是一种宝贵的科技信息源、商贸信息源和竞争情报源,是技术人员设计、制造新产品的一种有价值的参考资料,也是计划、开发、采购、销售及外贸等专业人员了解各厂商产品现状、掌握产品市场情况及发展动向的重要信息源。

10) 其他零散资料。其他零散资料主要指档案资料、舆图、图片和乐谱等零散资料。零散资料规格不同,外形各异,无法装订成册,通常需专门的方法进行整理和保管。

(2) 按文献载体形式划分

1) 印刷型文献。印刷型文献是以纸质材料为载体,以印刷(包括手写以及铅印、油印、石印、胶印和木版印刷等)为记录手段而产生的一种传统的文献形式。它具有悠久的历史,是传统的记录知识信息的方式。

2) 缩微型文献。缩微型文献又称缩微资料、缩微复制品,它是一种以感光材料为存储介质,以缩微照相为记录手段而产生的文献形式。

3) 视听型文献。视听型文献又称声像资料或直感资料。它是以磁性材料或感光材料为载体,利用声像技术和装置,直接记录声音、文字和图像的文献形式。

4) 机读文献。机读文献是将信息以计算机可读的形式,记录在磁性载体(磁盘、磁带等)、光学载体(光盘)上,在计算机软件的支持下读取的一种文献形式。

(3) 按文献加工层次分

1) 零次文献。零次文献指未经出版发行的或未以公开形式进入社会交流的最原始的文献。如私人笔记、底稿、手稿、考察记录、实验记录、会议记录、原始录音、原始录像、谈话记录、内部档案、论文草稿、设计草稿、原始统计数字、技术档案以及现代的 Email 通信、BBS 公告板文章、BLOG 和讨论组文章等。

2) 一次文献。一次文献是以作者本人的研究成果为基础而创作或撰写的文献。诸如图书、期刊论文、专利文献、会议文献、学位论文、技术档案、报纸、标准文献、图谱、研究报告、学位论文、政府出版物和产品样本等, 均属一次文献。

3) 二次文献。二次文献是对分散的、无组织的一次文献进行搜集、提炼、浓缩、加工及整理, 并按一定的科学方法组织编排、编辑出版的文献, 是为了便于管理和利用一次文献而编辑、出版的工具性文献。诸如目录、索引、文摘及检索引擎等均属二次文献。

4) 三次文献。三次文献是根据一定的目的和需求, 在大量利用一、二次文献资源的基础上, 对其内容进行综合分析、系统整理、高度浓缩及评述等深加工而形成的再生文献。如综述、述评、词典、百科全书、年鉴及书目之书目等都是三次文献。

总之, 现代文献的类型突破了纸质载体印刷品的传统模式, 采用多种载体材料记录文字、声音和图像来传播知识信息。在今后相当长的历史时期, 将形成各种文献形式长期共存、相互补充、相互结合、共同发展的局面。

3. 数字文献的概念

数字文献指以数字形式把文字、图像、声音和动画等多种形式的信息, 存储在光、磁等非纸介质的载体中, 并通过网络通信、计算机或终端等方式再现出来的资源。所谓数字形式, 是以能被计算机识别的、不同序列的“0”和“1”构成的形式。

数字文献的范围非常广泛, 其类型随着信息技术的迅速发展而日益丰富, 不同类型的数字文献的特点、功能存在一定的差异。依据不同的标准, 可以划分为不同的类型, 主要有以下几种类型。

(1) 按数字文献内容表现形式划分

可分为数据库、电子图书、电子期刊、电子报纸等。

1) 数据库, 是以一定的组织方式存储在一起的相关数据的集合, 是数字文献最早的形式, 也是数字文献的主要构成部分, 包括全文数据库、文摘/题录数据库和参考数据库等类型。

① 全文数据库, 即收录有原始文献全文的数据库, 以期刊论文、会议论文、政府出版物、研究报告、法律条文和案例以及商业信息等为主。如国内著名的全文数据库有中国知网的《中国期刊全文数据库》、重庆维普的《中文科技期刊数据库》等; 国外有 ProQuest 系统、EBSCOhost 系统以及 Springer Link 系统等。

② 文摘/题录数据库, 是指包含各种数据、信息或知识的原始来源和属性的数据库。它报道文献信息存在, 揭示文献信息内容。按数据库内容, 可划分为书目数据库、文摘数据库和索引(题录)数据库。书目数据库主要是针对图书进行内容及存储地址的报道与揭示的, 如各图书馆的馆藏机读目录数据库; 文摘和索引数据库则相对期刊论文、

会议论文、专利文献和学位论文等进行内容和属性的认识与加工,它提供确定的文献来源信息,供人们查阅和检索。科学引文索引(SCI)、社会科学引文索引(SSCI)、工程索引(EI)和科学技术会议录索引(ISTP)是世界上最权威的四大索引,均属文摘/题录数据库。

③ 参考数据库,是指包含大量事实、数据,以及其他类型文献信息的数据库,具体有数值数据库、指南数据库、术语数据库、视频数据库、英语学习数据库和考试系统等。数值数据库,指以数值为主要内容的数据库,如统计数据库、化学反应数据库等;指南数据库,是有关机构、人物等相关信息的简要描述,如公司名录、产品目录等;术语数据库,即专门存储名词术语信息、词语信息等的数据库,如电子版百科全书、网络词典等。视频数据库,是指能够以多媒体方式播放的相关信息内容的数据库,如爱迪科森网上报告厅、超星学术视频等;英语学习系统,主要是指以英语学习为主要内容的数据库,如新东方多媒体学习库、冰果英语和EBSCO的学生英语学习中心(SRC)等;考试系统是指提供试题和模拟考试的数据库系统,如银符考试系统、起点考试系统等。

2) 电子图书是指直接以数字形式出版的图书或印刷型图书的数字化形式,是以电子版的方式(其载体主要为光盘或网络等)呈现的,并通过计算机或电子图书阅读器进行阅读的一种电子文献。如国外的NetLibrary、Ebrary;国内的超星电子图书、书生电子图书和方正Apabi电子图书等。

3) 电子期刊,就其广义而言,是指所有以电子形式存在的期刊,包括印刷型期刊的电子版、无印刷版的网络型电子期刊。严格意义上的电子期刊是纯电子期刊,即从投稿、编辑出版、发行订购到阅读的全过程都是在网络环境中进行的,任何阶段都不需要用纸,因而与传统的印刷型期刊有着本质的区别。与纸本期刊并行的电子期刊,如著名的《科学》(Science)、《自然》(Nature)、Elsevier/Wiley/Springer/Kluwer等出版商的电子期刊等;纯电子期刊,如《数字图书馆杂志》(D-Lib Magazine)。

4) 电子报纸,指综合运用多媒体技术、网络技术和通信技术,完成出版、发行、利用全过程的新闻媒体。电子报纸的内容可以是文字、表格、彩色图形、图像,甚至是声音、动画等的多媒体信息,用户可以在任何地点利用计算机网络进行信息检索、存储和利用。

(2) 按数字文献性质和功能划分

可分为一次数字文献、二次数字文献和三次数字文献。

1) 一次数字文献,指反映最原始思想、成果、过程以及对其进行分析、综合、总结的信息资源,如电子图书、电子期刊、电子报纸、事实数据库以及发布原始信息的学术网站等。用户可从一次数字文献中直接获取自己所需要的原始信息。

2) 二次数字文献,指对原始信息进行加工、整理,便于利用原始文献的信息资源,如参考数据库、网络资源学科导航和搜索引擎/分类指南等。二次数字文献是检索一次数字文献的有效工具。

3) 三次数字文献,指对二次数字文献进行综合分析、加工和整理的信息资源,如专门用于检索搜索引擎的搜索工具,比较典型的是WebCrawler。

(3) 按数字文献生产途径和发布范围划分

可分为商用电子资源、网络公开学术资源和特色资源。

1) 商用电子资源, 亦称正式电子出版物, 由正式出版机构或出版商 / 数据库商出版发行, 在数字学术信息资源中所占比例最大, 包括各类数据库、电子期刊和电子图书。其特点: 学术信息含量高, 具备检索系统, 便于检索利用, 出版成本高, 必须购买使用权才可以使用, 因此并不是面向社会公众免费开放的。

2) 网络公开学术资源, 是完全面向公众开放使用, 包括各种学术团体、行业协会、政府机构、商业部门和教育机构等在网上正式发布的网页及其信息, 用于检索图书馆印刷型馆藏的联机公共目录系统 (OPAC) 也属于这部分范畴。

3) 特色资源, 主要基于各教育机构、政府机关和图书馆的一些特色收藏制作, 在一定范围内分不同层次发行, 不完全面向公众。如学位论文、教学课件 (CAI), 一般只在校园网内作为教学资料使用。

(4) 按数字文献生产方式划分

可分为原生数字文献和复合数字文献。

1) 原生数字文献是通过科研人员有意识的思维活动, 把客观存在的离散的原生态信息源经过一定的积累、分析、比较、判断、归纳和重组, 经过数字化处理, 创造出的符合客观实际的、知识化的数字文献。由于原生数字文献包括已公开发表的一次文献资源和未公开发表的灰色文献资源, 故原生数字文献具有原始性和原创性等特征。

2) 复合数字文献多数是从其他载体 (如纸介质) 文献转化而来的数字文献。这在当前及今后一个时期, 仍然是我国图书馆馆藏数字文献建设的主要形式之一, 但随着社会信息化程度的提高, 原生数字文献的增长速度在加快。

(5) 按数字文献数据传播范围划分

可分为网络信息资源和单机信息资源。

网络信息资源是指借助于计算机网络可以获取和利用的数字文献的总和; 单机信息资源是指通过计算机存储和阅读但不在网络上传输的数字文献。两者的区别就在于其存储和传播的空间范围。

(6) 按数字文献存储介质划分

可分为磁介质和光介质。

磁介质包括磁盘、磁带、磁盘阵列、硬盘和活动硬盘等。随着信息存储技术的发展, 超大容量的外存储系统磁盘阵列以及可移动磁盘相继问世, 成为重要的数字信息载体。光介质是一种用激光来记录和再现信息的高密度存储介质, 即光盘。

4. 数字文献的特点

数字文献在生产制作方式、信息符号、载体材料以及载体形态等方面与传统的纸质文献完全不同, 因此它也具有与传统纸质文献完全不同的特点。

1) 高度的共享性。数字文献是以现代信息技术为生产手段而形成的一种数字化的信息, 其共享性主要体现在两个方面: 第一, 数字信息可以进行大量的无差别复制, 其复制品不仅不存在任何质量和功能上的缺陷, 而且复制的成本极为低廉, 并对信息源没有任何损害, 数字文献从本质上来说就是一种易于共享的资源。第二, 数字化的信息以

机读数据的形式存在,可以通过通信网络进行远距离传送,不受时间和空间的限制,这就使数字文献在全球的共享成为现实。

2) 类型的多样性。数字文献包含的信息类型多种多样,从电子报刊、电子工具书、商业信息、新闻报道、书目数据库、文献信息索引,到统计数据、图表、电子地图等,既可以是文字、图表等静态信息,也可以是集图、文、声、像于一体的动态多媒体信息,并且各种类型的数据又可借助计算机实现任意的组合编辑,具有打印输出及磁盘输出等多种输出方式。这就使得信息的组织方式发生了巨大的变化,不仅以知识和信息为基本单元,而且充分展示这些单元间的逻辑关系。

3) 检索的便利性。数字文献将文字、图像及声音等变成磁性介质上的电磁信号或光介质上的光信息来进行信息的存储和传递,具有超强的检索功能,检索信息的速度快、途径多,且不受时间地点的限制。特别是网络信息资源检索利用超文本链接,构成立体网状文献链,能把不同地点的文献都通过节点链接起来,使人们能够在信息海洋中快捷、准确地得到所需的有关信息。

4) 信息的时效性。数字信息从根本上改变了信息的创造、交流和获取的方式,无纸化的出版从作者投稿、专家审稿到稿件编辑等都在网上进行,大大缩短了文献的编辑出版时间,从而使信息的时效性大大增强。而且,数字文献的更新也比传统出版物的修订容易得多,可以随时修改内容,及时将信息发布给用户。

5) 信息的不安全性。数字信息从形成、传输到存储都是通过计算机或具有类似功能的设备实现的,所以数字信息对计算机系统各种软硬件设备具有很强的依赖性,数字信息的长期保存面临着严峻的挑战。在网络环境下,数字信息安全还受到计算机病毒、黑客攻击等威胁。数字信息的易复制性,也给数字信息的知识产权保护带来诸多问题。

1.1.2 数字文献的检索语言

1. 数字文献检索

数字文献检索是指通过检索系统,采用一定的技术手段,根据一定的原则,在数据库或其他形式的数字文献中自动找出用户所需相关信息的过程,具体如图 1.1 所示。广义的检索包括信息的存储和检索两个过程,狭义的检索仅包括用户的信息检索过程。在数字文献检索发展过程中,先后经历脱机检索、联机检索、光盘检索和网络检索四个阶段。目前,网络检索已成为广大用户获取信息的主要方式。

1) 脱机检索:以磁带为存储介质,是早期的数字文献检索形式,专职检索员对用

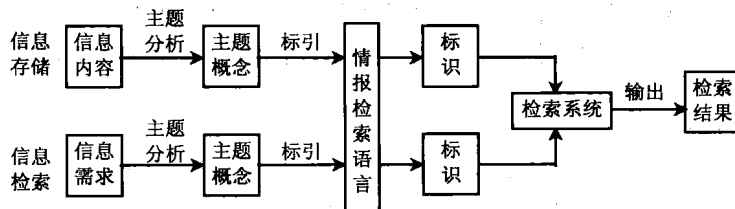


图 1.1 广义的信息检索过程

户的检索要求进行批处理检索, 人机之间没有交互性。

2) 联机检索: 以联机检索提供商为中心, 提供商研制自己的检索软件, 建立自己的联机检索系统, 数据库则是从数据库生产商那里购买的。用户利用联机检索终端, 通过专用的或公用的网络线路等数据通信网络与联机检索系统相连, 按照提供商所制定的各项检索规则进行检索。由于互联网技术的发展, 联机检索系统的可操作性大大增强。

3) 光盘检索: 以光盘为存储介质, 从单机光盘到光盘塔、光盘库的局域网共享。但随着磁盘容量的不断扩大, 光盘检索逐渐被淘汰。

4) 网络检索: 数据可以集中存储在某一台主机上, 也可以分散存储于网络中各个主机上。对用户而言, 只要用户的电脑能上网, 并在得到数据库的使用授权后, 便可以通过网络访问远程数据库或建立在本地局域网内的镜像数据库。

2. 数字文献检索语言

检索语言, 是信息存储与检索过程中用于描述文献特征和表达用户信息提问的一种专门语言。所谓检索的运算匹配就是通过检索语言的匹配来实现的。

检索语言是人与检索系统对话的基础。有了这种语言, 标引人员在进行信息存储的过程中, 就会对原始信息进行分析, 找出其能代表信息的特征与检索语言(检索标识系统)进行对照标引, 然后纳入检索系统; 而检索人员在进行文献检索的过程中, 则先对待查课题进行分析, 归纳出各种信息特征, 使之形成能代表需要的检索提问, 然后把这些提问与检索语言(检索标识系统)进行核对, 标引成检索提问标识。

如果没有检索语言作为标引人员和检索人员的共同语言, 就很难使标引人员对文献信息内容的表达和检索人员对相同文献信息内容需求时的表达相一致, 信息检索也就不可能顺利实现。图 1.1 也同时展示了检索语言在信息存储和检索过程中的作用。

文献具有外部特征和内容特征。外部特征有题名(书名、篇名和刊名)、著者(个人、团体)、号码(专利号、标准号、国际标准书号 ISBN 和国际标准刊号 ISSN 等)、机构名(责任者所属机构、出版发表机构)以及时间等; 内部特征指文献内容所属的学科和文献内容的研究对象。

随着计算机检索的极大应用, 对文献的外部和内容特征进行著录与标引的检索语言主要是分类检索语言和主题检索语言, 其中主题检索语言中目前应用最广泛的是关键词检索语言。描述文献内容所属学科性质的检索语言为分类检索语言, 描述文献研究对象的检索语言为关键词检索语言。体现文献内容学科属性的分类常用在分类检索途径(检索字段)或限制学科检索范围上。体现文献内容研究对象的关键词常用在题名、关键词和文摘(摘要)检索途径上, 甚至是全文检索途径上。而一些外部特征, 如文献的责任者、号码、机构(作者所属单位)和年月等也可以作为检索途径。文献检索语言与检索途径的关系见图 1.2 所示。

(1) 分类检索语言

分类, 是人类认识和把握事物、认识和把握知识的基本思维方法。事物、知识经过分类后就能全面地从不同角度显现其内在的层次与联系。分类检索语言是用分类号来表达各种概念, 将各种概念按学科性质进行分类和系统排列的检索语言。使用这种检索语言建立的信息检索系统可以反映知识的从属、派生、重合、交叉和并列等关系, 用户因

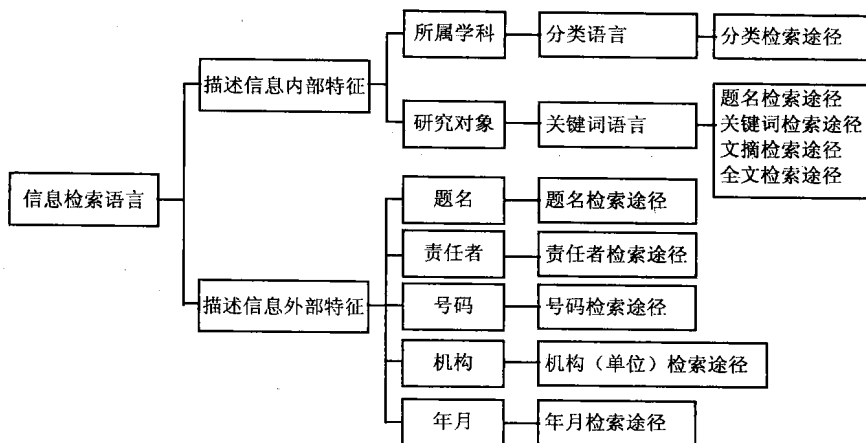


图 1.2 文献特征与检索途径

此可以系统地掌握和利用一个学科或专业范围的信息。分类检索语言包括体系分类法和组配分类法，目前数字文献检索采用的大多为体系分类法。

体系分类法是指将分类表中的全部类目列举出来，并按照特定的等级系统将信息组织起来的一种分类方式。这种分类法的基本特点：一是对知识进行逻辑的、系统的组织，因此称为体系分类法或系统分类法；二是把知识分门别类地进行等级排列，因此也称为等级分类法；三是把知识详细地罗列、展开，因此又称为列举分类法或枚举分类法。

体系分类法的目的就是编制或建立一个知识地图、一个知识树，既囊括人类的全部知识领域，又分门别类地进行聚类、梳理和组织，使用户可以从知识的整体结构或其中的一个知识节点为入口，层层深入，检索到相关的信息、知识和文献。

体系分类法主要形成按学科或专业集中文献的严密的知识概念等级系统，体现了学科系统性，便于从学科或专业的角度出发进行族性检索。在这个系统中知识按其内容的亲疏远近有序地排列，检索者可以查找和利用某一个专业范围的知识，对从学科或专业出发的检索课题可达到相当高的检全率，这是其他检索系统所无法达到的。例如检索“中国高等教育”的文献，在一个类目下几乎可以检全全部文献。如果浏览的初始入口选择得比较准确，检准率也很高。

传统的体系分类法由于其固有的等级逻辑结构、线性体系以及针对手工检索工具的标引规则等原因，导致它在很多方面检索效率不高，主要表现如下。

- 1) 体系分类法受分类层次的限制不需要也不可能详尽地划分和列举，导致检索专指度受到限制，特别是隐含的知识主题不经过反复地浏览不易检索到。
- 2) 体系分类法“分散与集中”的矛盾，使检全关于某种事物的文献信息十分困难。
- 3) 体系分类法按学科和专业集中文献信息以及线性的分类体系，使其在处理学科之间互相交叉渗透和综合而形成的新知识领域很困难，检索效率低。
- 4) 在体系分类法先组定组式的标记系统中，分类号是一个不可拆分的整体，受到分类标记制度和标引规则的限制，难以实现有效的组配检索。

目前，国内使用最广泛的分类法是 1999 年版的《中国图书馆分类法》（第四版），