

SHIYONG KEJIXINXIZIYUAN JIANSUOYULIYONG
SHIYONG KEJIXINXIZIYUAN JIANSUOYULIYONG
SHIYONG KEJIXINXIZIYUAN JIANSUOYULIYONG
SHIYONG KEJIXINXIZIYUAN JIANSUOYULIYONG

SHIYONG KEJIXINXIZIYUAN JIANSUOYULIYONG

实用科技信息资源 检索与利用 (第三版)

主 编 马光雄 袁丽芬 王苏海
副主编 颜 惠 冯 进 徐雪梅

SHIYONG KEJIXINXIZIYUAN JIANSUOYULIYONG
SHIYONG KEJIXINXIZIYUAN JIANSUOYULIYONG
SHIYONG KEJIXINXIZIYUAN JIANSUOYULIYONG
SHIYONG KEJIXINXIZIYUAN JIANSUOYULIYONG

IMAGE
MORE

南京
大学
出版
社

实用科技信息资源 检索与利用 (第三版)

主 编 马光雄 袁丽芬 王苏海
副主编 颜 惠 冯 进 徐雪梅
编写人员 马光雄 袁丽芬 王苏海 颜 惠
冯 进 徐雪梅 张 鑫

SHIYONGKEJIXINXIZIYUAN JIANSUOYULIYONG
SHIYONGKEJIXINXIZIYUAN JIANSUOYULIYONG
SHIYONGKEJIXINXIZIYUAN JIANSUOYULIYONG
SHIYONGKEJIXINXIZIYUAN JIANSUOYULIYONG



南京大学出版社

内容提要

本书为有关科技信息资源检索与利用方法的实用教程。全书以理工科大学生为对象,以培养信息意识、检索理念和检索技能为目的,以信息检索策略为主线,涉及检索原理、互联网信息资源、中外文科技文献检索、特种文献及其检索以及科技论文写作等内容。其第一版曾于2009年荣获华东地区大学出版社第八届优秀教材学术专著二等奖及中国大学出版社首届优秀教材二等奖。为紧跟信息检索系统及检索技术的发展和变化,本书曾于2010年推出第二版,并增添了文献信息资源免费共享一章。本次第三版除内容全面刷新以外,还增添了引文检索、学术指标分析、RSS订阅服务、文献管理软件 NoteExpress、文献分析软件 RefViz 等内容,因此本版除可供本科生使用外,也可供研究生复习和提高之用。此外,由于书中介绍了收费数据库的免费使用方法,因此还特别适合工程技术人员使用。

图书在版编目(CIP)数据

实用科技信息资源检索与利用 / 马光雄, 袁丽芬,
王苏海主编. — 3版. — 南京: 南京大学出版社,
2013. 8

ISBN 978-7-305-04988-0

I. ①实… II. ①马… ②袁… ③王… III. ①科技情
报—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(2013)第 047641 号

出版发行 南京大学出版社

社 址 南京市汉口路 22 号 邮 编 210093

网 址 <http://www.NjupCo.com>

出版人 左 健

书 名 实用科技信息资源检索与利用(第三版)

主 编 马光雄 袁丽芬 王苏海

责任编辑 吴 华 编辑热线 025-83596997

照 排 南京南琳图文制作有限公司

印 刷 常州市武进第三印刷有限公司

开 本 787×1092 1/16 印张 15 字数 384 千

版 次 2013 年 8 月第 3 版 2013 年 8 月第 1 次印刷

印 数 1~3600

ISBN 978-7-305-04988-0

定 价 29.80 元

发行热线 025-83594756

电子邮箱 Press@NjupCo.com

Sales@NjupCo.com(市场部)

* 版权所有,侵权必究

* 凡购买南大版图书,如有印装质量问题,请与所购
图书销售部门联系调换

序

就在2007年新年的钟声刚刚敲响的时候,我读到了江苏科技大学图书馆的老师们即将付印的教材《实用科技信息资源检索和利用》的样书,眼前豁然一亮。我与本书的一些作者曾经是同事,看到他们精心编写的教材,就像看到他们精心护理的婴儿终于长大一样,由衷地为他们感到高兴与骄傲。

随着社会信息化进程的不断加快,科学技术的日新月异和信息处理技术的突飞猛进,工程技术信息媒体的形式及其传递方式已经或者正在发生着巨大的变化。网络信息的飞速增长极大地改变了人们获取信息的方式,因特网已经成为我们学习、工作与生活中不可分割的重要部分。随着因特网技术的普及,近年来网络工程技术信息资源与日俱增,网络环境给工程技术人员展现了一个五彩缤纷的技术信息世界。然而,我们不得不看到另外一种情形:传统的信息检索理论和方法已经出现了很大的变化,在网络化和数字化时代,人们获取网络信息的能力与迅速发展的网络信息资源极不相称。数字化技术、全文检索技术、因特网技术、多媒体技术和软硬件的频繁更新换代使得即使有一定水平的专业人员也难以停留在原有知识的基础上。在网络信息的海洋中冲浪,需要不断更换新的工具、采用新的方法、获取新的知识和技能。面对浩如烟海却又纷繁复杂的网络工程技术信息资源,一个普通的工程信息用户要想全面准确地搜寻对自己有用的信息就会感到越来越迷茫,越来越无助。他们首先要解决的问题是:科学技术信息资源主要有哪些?这些资源在哪里?如何全面、准确、快速地检索到特定的信息?江苏科技大学图书馆的老师们在这本书中就针对这些问题作出了全面的、系统的并且深入浅出的解答。

综观全书,人们不难看出,这本书从头至尾贯穿了一条主线,这就是结合理工科院校人才培养的具体需要,结合科技信息检索课程教学的实际,还结合了理工大学图书馆所拥有的网络信息资源,体现了实用性这一特色。全书针对非图书情报专业教学对象而设计,不拘泥于文献的类型,无论是印刷版工具,还是馆藏数据库,或者是网络免费资源,只要与本书的宗旨相关,则都在本书收录与讨论之列。这本书中所介绍的信息检索工具和相应的运用方法都是高效率获取各类信息所应该掌握的。每个章节的作者都紧紧围绕着工程技术信息资源选材,这就在很大程度上拉近了作者与读者(这些读者中更多的是工程技术人员以及准技术人员)之间的距离。同时,该书的新颖性也十分明显。无论是信息资源的推荐,还是国外技术的介绍,作者都以最新的眼光去观察、去取材,许多章节都给人们以耳目一新的感觉。另外,书中大量的实例、练习题、思考题,反映了作者大量的信息检索的实践和实战的基础。只有凭借大量的检索实践操作的积累,并且始于实践又

不止于实践的人,才能成为一个优秀的信息检索教员的候选人。而建立在操作实践基础之上的教材,也就更容易学习,更便于推广。还有,全书文笔简练,朴实无华,毫无造作痕迹,使人们读之很有亲切感,不易产生阅读疲劳,这也在很大程度上降低了读者阅读与自学的难度。相信这本书的出版,无论是对江苏科技大学的学生的信息检索课程的教学,还是对工程技术人员熟悉科技领域的信息资源,都是一种有意义的贡献。

自从国家教育部文件《关于在高等学校开设“文献检索与利用”课程的通知》发布后,二十多年来,我国高等院校的信息检索课程有了较快的发展。这对培养学生的自学能力、获取信息的能力、动手能力和创新能力,具有积极的作用。江苏科技大学自1985年春季学期起由图书馆负责向全校开设文献检索类课程,是当时国内较早开设这门课程的高校之一。多年来,他们结合本校具体的需求,结合本校的实际个案开设信息检索选修课、必修课,使得具有船舶工程技术特征的信息检索课程长盛不衰,使得数万名工程技术专业的大学生接受了信息素质的基本培训。如今,有了这本结合该校具体需求、结合该校科研教学方面具体实例的教材后,相信这门课程在江苏科技大学会得到更多的机会,取得更大的发展,获得更多的成果,并赢得教师与学生更大程度的认可与赞誉。

华薇娜

2007年元月于南京大学

三版前言

2007年出版的《实用科技信息资源检索与利用》一书,自2010年2月再版以来,又已三年有余,虽然时间还不算太长,但国内外检索系统的面貌,却发生了很大变化,为了能与时俱进,我们组织相关老师进行了第3版的修订工作。借修订之机,我们也向使用本教材的各校同仁,征求了对本书的意见和修改建议,冀求能最大程度地满足各校师生的需要。

本书第3版沿用原2版的结构,但在内容上有所创新,其主要特点如下:

(1) 彻底重写了原第1章信息检索基础知识的内容,详细介绍了三种数据库(数据型数据库、文献数据库、网页数据库)的结构;分别阐述了5种索引(整字段索引、英文全文索引、中文全文词索引、中文全文字索引、数值索引)的原理;突出了检索语法的概念,强调了项间逻辑算符和项内逻辑算符的区别。希望能在此基础上,适当提高教材的深度,帮助读者更好地领会各种检索系统的语法规则。

(2) 彻底重写了原第2章互联网信息资源搜索的内容,详细阐述了搜索引擎的高级搜索语法,补写了网上参考资源一节,引入了网络个性化信息平台的概念。

(3) 彻底重写了原第3章国内文献信息检索系统的内容,按照国内文献信息检索系统的最新面貌,详细介绍了维普、万方、CNKI的新工作界面和检索语法,以及引文分析、学术指标分析等最新功能。

(4) 认真修订了第4章国外文献信息检索系统的内容,按照国外文献信息检索系统的最新面貌进行了修订,并从实用出发,删去了原CA on CD一节,补写了IEEE电子图书馆,即IEEEXplore一节。

(5) 认真修订了第5章特种文献及其检索系统的内容,补充了关于优先权项、日本专利的介绍,并将原第7章中的文献辨识一节移入本章,重点介绍参考文献表中的外文特种文献的辨识方法。

(6) 近年来国内公共图书馆在推进文献资源免费共享方面,取得了很大进展,我们尽量将这些变化收入新版第6章之中,以给工程硕士读者更多帮助。

(7) 原第7章信息资源综合利用也做了较大程度的改动。随着各种文献管理和文献分析软件的出现,传统的手工操作多已可通过计算机来实现了,为此,我们删去了原2版的部分内容,换之为:利用NoteExpress软件进行文献管理、利用

RSS 订阅服务功能进行文献跟踪、利用 RefViz 软件进行文献分析等 3 节。

(8) 由于新版增添了有关高级检索语法、引文检索分析、学术指标分析、计算机辅助文献管理、计算机辅助文献分析、RSS 文献订阅推送等方面的内容,因此能更好地满足研究生读者的需要了,教师可根据教学对象,灵活组织教学内容。

(9) 为进一步提高本教材的可读性和趣味性,对于原来排在各章末尾的补白部分,增添了内容,分别插入有关章节之中,并给予小贴士、想一想、趣闻等小标题。

参加第 3 版修订工作的老师有:

第 1 章	信息检索基础知识	马光雄		
第 2 章	互联网信息资源搜索	袁丽芬	徐雪梅	马光雄
第 3 章	国内文献信息检索系统	冯 进	马光雄	
第 4 章	国外文献信息检索系统	颜 惠	马光雄	
第 5 章	特种文献及其检索系统	王苏海	马光雄	
第 6 章	文献信息资源的免费共享	马光雄	袁丽芬	
第 7 章	信息资源综合利用	袁丽芬	张 鑫	马光雄

本书由马光雄负责主持改版、全面润色和最终定稿,袁丽芬负责全面审校,王苏海负责组织和协调工作。

在此次修订工作中,得到了江苏科技大学图书馆领导及同仁们的关心和支持,苏州制氧机厂徐桢基高级工程师也积极提出修改建议并提供检索实例,在此一并向他们表示感谢。

我们也要感谢对本书历次改版工作给予帮助的南京大学出版社,是你们的关心、支持和推动,才使得本书能从 2007 年顺利地走到今天。

最后还要感谢采用本教材的兄弟院校师生连续多年的扶持,是你们一如既往的关心和指正,才使得本书能不断地改进和成长。

编 者

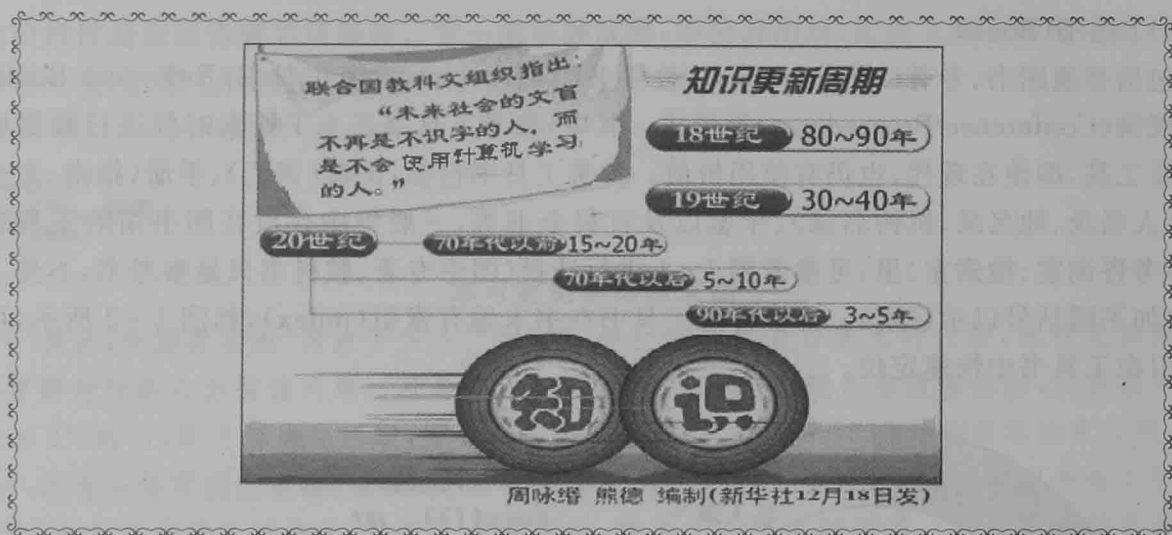
2013 年 1 月

目 录

第 1 章 信息检索基础知识	1
1.1 信息检索原理	1
1.2 信息检索语言	12
1.3 信息检索语法	14
1.4 信息检索策略	21
1.5 电子文献阅读	26
练习与思考 1	27
第 2 章 互联网信息资源搜索	29
2.1 网络信息搜索概述	29
2.2 网络信息高级搜索	30
2.3 网络学术文献搜索	36
2.4 网络参考信息资源	40
2.5 网络个性化信息平台	47
练习与思考 2	49
第 3 章 国内文献信息检索系统	50
3.1 读秀学术搜索	50
3.2 维普期刊资源整合服务平台	52
3.3 万方数据知识服务平台	65
3.4 中国知识资源总库 CNKI	80
练习与思考 3	91
第 4 章 国外文献信息检索系统	92
4.1 Engineering Village	92
4.2 ScienceDirect	106
4.3 EBSCO host	114
4.4 SpringerLink	123
4.5 IEEEExplore DIGITAL LIBRARY	127
练习与思考 4	133
第 5 章 特种文献及其检索系统	135
5.1 专利文献及其检索系统	135
5.2 标准文献及其检索系统	153
5.3 会议文献及其检索系统	158
5.4 学位论文及其检索系统	160

5.5	科技报告及其检索系统	163
5.6	外文特种文献辨识	168
	练习与思考 5	170
第 6 章	文献信息资源的免费共享	172
6.1	公共图书馆数字资源免费共享	172
6.2	部分省市的科技文献资源共享平台	186
6.3	开放存取学术资源	188
	练习与思考 6	193
第 7 章	信息资源综合利用	194
7.1	利用 NoteExpress 软件进行文献管理	194
7.2	利用 RSS 订阅功能进行文献跟踪	205
7.3	利用 RefViz 软件进行文献分析	213
7.4	通过科技写作实现信息资源增殖	220
7.5	遵循国家标准写出规范的学位论文	224
	参考文献	231

第 1 章 信息检索基础知识



1.1 信息检索原理

1.1.1 信息检索的概念

1. 什么是信息检索?

信息检索能力是信息素质的重要构成成分,也是当代大学生必须具备的学习能力。如果不会使用计算机进行信息检索,就有可能陷入新文盲的尴尬境地哦!

那么,什么是信息检索(information retrieval)呢?检索(retrieval)、搜索(searching)是一回事吗?虽然检索与搜索常被等同混用,但严格地说,只有在已经有序存储的信息集合中进行搜索才是信息检索,如果是在杂乱的信息资料中随机查找,则仍然只是搜索,因此,用户如能知道一些信息存储的原理,对提高检索效率是会大有帮助的。

2. 信息检索大家族

现代信息检索已计算机化和网络化,按照所检信息的特征,一般可区分为:特定信息检索和信息资源检索两大类。特定信息检索又称为数据检索,它是一种确定性的检索,包括事实检索和数值检索,检索后立即获得所需的信息(事实或数值)。信息资源检索则是一种相关性的检索,其检索目标是与所需信息相关的信息资源。所谓信息资源即可被人们利用的信息集合,包括口传信息资源、实物信息资源、文献信息资源、网络信息资源等,常见的文献检索或网页搜

索就属于信息资源检索的范畴,检索后,需阅读该资源的原文,才能获知其中所含的信息。

网页搜索的工具叫做搜索引擎。事实检索、数值检索、文献检索三者的工具则都是专门的数据库,故习惯上又把这三者统称为数据库检索,其中文献检索是本书讨论的重点。

1.1.2 文献信息资源

1. 科技文献的类型

科技界常把文献按照出版形式划分为以下十大类型:

(1) 图书(Book)

包括普通图书、专著(Monographs)、教科书(Textbook)、参考工具书(Reference Book)以及会议录(Conference Proceedings)等形式。其中,参考工具书是手工检索时代进行数据检索的主要工具,即使在现代,也仍有使用价值。参考工具书包括:词典(词汇)、手册(指南、大全)、名录(人名录、地名录、机构名录)、年鉴以及百科全书等,一般集中存放在图书馆的工具书室(或参考咨询室、检索室)里,可参考图 1-1 进行选择(图中专著、教科书只是参考书,不是工具书,故加了圆括号以示区别)。许多参考工具书在书末编有索引(index),如图 1-2 所示,可利用索引在工具书中快速定位。



图 1-1 参考工具书选择方法示意图

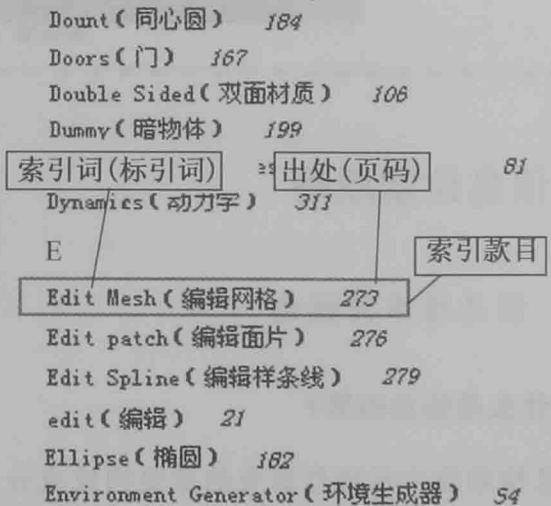


图 1-2 《3D Studio MAX 速查手册》的书末索引

(2) 报刊

包括报纸(Newspaper)和期刊(Journal, Periodical, Serial),其中的科技期刊还可分为:学术性期刊、技术性期刊和科普性期刊。如果在刊名中可以找到高校(University, College)、科研院所(Institute, Institution)、专业学会(Society, Institute, Institution)等机构名称,并且可以找到“学报”(Acta)、“会刊”(Proceedings)、“会报”(Bulletin)、“汇刊”(Transactions)、“评论”(Review)等字样,则多半为学术性期刊。如果刊名只是某一技术领域名称(如 Welding Journal),则多半为技术性期刊。学术性期刊和技术性期刊是科技人员经常浏览和进行检索的对象,应特别加以关注。

(3) 会议论文(Conference Paper)

指在各种学术会议上发表的论文,它们有的出现在以图书形式出版的会议录(Proceed-

ings)里,有的出现在以期刊形式出版的会议专辑里。会议论文的学术水平较高,而且发表及时,新颖性较强,因此和期刊论文一样具有较高的信息价值。

(4) 学位论文(American多称为 Dissertation,英国多称为 Thesis)

包括学士论文、硕士论文和博士论文,但作为传播、交流和检索对象的主要是后两种。学位论文不公开出版,只存放在学位授予单位和指定收藏单位中,如我国的国家图书馆和中国科技信息研究所。

(5) 科技报告(Technical Report)

科技报告是科研人员从事专题研究取得部分或全部研究成果后,向科研资金资助单位提交的阶段性进展报告或总结报告。它一般单独成册,不公开出版,但经上级主管部门统一编号后可对外交流(军工科技报告需在解密后)。



趣闻

本可避免的昂贵支出

据说,美国在实施“阿波罗登月计划”中,曾付出数以百万美元的代价,来研究如何解决甲醇对钛的应力腐蚀问题。可是事后查明,早在10年前,就已有报告指出,只要在甲醇中加2%的水,即可解决此问题,而要获得这一信息,只需花费15分钟的信息检索时间。

还有一件可笑的事情,据说,美国有一家轧钢厂中的化学家,对该厂的图书馆工作人员讲,他们花费了1万美元,进行了大量试验,解决了一个关键问题。可是图书馆员却对他说,馆中正好有一份德国人的研究报告,研究的也是同样的问题,并且得到了相似的结论,而获取该报告,只需5美元。

(6) 专利文献(Patent Literature)

包括专利说明书、专利局公报、专利证书等,它是国家实行专利制度下的产物,是重要的而且易于获得的科技信息资源。

(7) 标准文献(Standard Literature)

包括技术标准、检验规程、建设规范或建造规范等,它是国家实行标准化制度的产物。

(8) 产品资料(Product Literature)

包括产品目录、产品样本、产品说明书以及厂商情况介绍等宣传品,它们多由厂商自行印发,印刷精美、内容生动、数据比较可靠,但新陈代谢频繁、时效性强,同时也比较零散,难以收集齐全,故一般只作为信息来源,不作为论文的参考引用文献。

(9) 政府出版物(Government Publication)

是由政府出版机构(如美国的政府出版局、英国的皇家出版局等)出版的文件、公告、法规、报告等资料,具有权威性和正式性,可从中了解到国家的方针政策、经济形势、科学技术发展水平等信息。我国政府需向社会发布的文件、法规等资料,由指定的出版社(如人民出版社、法律出版社等)出版发行。

(10) 档案(Archives)

是政府机构、企事业单位或个人在从事各种具体工作中积累下来的原始文件、图纸资料和真实记录,有重要的参考借鉴价值,但一般只在一定的范围内向外界提供借阅,故获取困难。

此外,从实用的角度来说,亦可将文献简单地划分为图书、期刊和非书非刊的特种文献三

大类;而从科技信息检索的角度来说,最常用的则是科技论文(包括期刊论文、会议论文和学位论文)、科技报告、专利说明书和技术标准 4 类文献。

从获取的难度来考虑,图书、期刊、会议录、汇编、专利公报、技术标准等公开发布、易于获得的文献,又可称为白色文献;学位论文、研究报告等内部发行或交流、较难获得的文献,称为灰色文献;处于保密状态的军工科技报告等从外部不能获取的文献,则称为黑色文献。利用现代的信息资源检索手段,可以获取大部分的白色和灰色文献,而要获取军工科技报告,一般要通过军工系统的内部专线,采用保密的检索手段,才能奏效。

最后,从文献的生成次序考虑,还可有一次、二次和三次文献之分。图书、期刊论文、会议论文等原创性文献被称为一次文献;图书目录、期刊论文索引等书本式检索工具被叫做二次文献;利用二次文献帮助,收集大量一次文献,综合写成的专题综述,或是手册、名录等参考工具书,则称为三次文献。

依此类推,又可划分出零次文献和高次文献。可想而知,零次文献就是还未进入信息交流领域的手稿、文稿、实验记录等原始文字;而高次文献则是在大量的一、二、三次文献基础上,通过信息分析、知识挖掘、知识发现等现代分析研究手段,生成的知识品位更高、知识内涵更深的新文献,例如,各种信息分析研究报告、可行性分析论证报告、新技术发展前景预测报告等。

2. 文献资源保障体系

文献信息资源是最宝贵的知识性信息资源,为保证文献资源的传播、交流、检索和传承,世界各国都高度重视文献资源的保存,建立文献资源保障体系。

我国的文献资源保障体系主要由各级图书馆(文献中心、信息中心)组成,拥有公共图书馆、高校图书馆和专业图书馆(如科学院图书馆及各行业信息研究机构中的文献馆)三大分支,我国部分文献资源保障单位见表 1-1 所示。

表 1-1 我国部分文献资源保障单位的通信地址、邮编、电话及网址

单位名称	通信地址, 邮编, 电话, 网址
中国国家图书馆·中国国家数字图书馆	北京市中关村南大街 33 号, 100081, 88545426, www.nlc.gov.cn
中国科学院国家科学图书馆	北京市中关村北四环西路 33 号, 100080, 82623303, www.las.ac.cn
中国社会科学院图书馆	北京市建国门内大街 5 号, 100732, 85195222, www.lib.cass.org.cn
国家农业图书馆·中国农业科技文献与信息服务平台	北京市中关村南大街 12 号, 100081, 68919892, www.nais.net.cn
上海图书馆·上海科技情报研究所	上海市淮海中路 1555 号, 200031, 64451207, www.library.sh.cn
天津图书馆	天津市南开区复康路 15 号, 300191, 23620082, www.tjl.tj.cn
重庆图书馆	重庆市沙坪坝区凤天大道 106 号, 400037, 65210822, www.cqlib.cn
南京图书馆	南京市中山东路 189 号, 210002, 84356000, www.jslib.org.cn
湖南图书馆	长沙市韶山北路 169 号, 410011, 84174121, www.library.hn.cn
四川省图书馆	成都市总府路 6 号, 610016, 86713056, www.sclib.org
广东省立中山图书馆	广州市文明路 213 号, 510110, 83810637, www.zslib.com.cn
中国科学技术信息研究所·国家工程技术数字图书馆	北京市复兴路 15 号, 100038, 58882584, www.istic.ac.cn/

续表

单位名称	通信地址, 邮编, 电话, 网址
中国国防科技信息中心(北京文献服务处)	北京市阜成路 26 号, 100036, 66358416, http://www.cetin.net.cn/cdstic/
机械工业信息研究院·全球机械文献资源网	北京市百万庄大街 22 号, 100037, 68993980, www.gmachineinfo.com/
冶金工业信息标准研究院·冶金信息网	北京市灯市口大街 74 号, 100730, 65262739, www.metalinfo.com.cn
中国化工信息中心	北京市安定门外小关街 53 号, 100029, 64444120, www.cncic.gov.cn
电子科学技术情报研究所·科技文献与数据中心	北京市石景山区鲁谷路 35 号, 100040, 68667671, www.eisc.com.cn
中国船舶信息网络中心·中国船舶在线	北京德胜门外双泉堡甲 2 号, 100085, 64872211, www.shipol.com.cn
中国核科技信息与经济研究院·中国核信息网	北京市海淀区阜成路 43 号, 100048, 68410207, www.atominfo.com.cn
中国专利信息中心	北京市蓟门桥西土城路 6 号, 100088, 62056956, www.cnpat.com.cn
国家标准馆·国家标准文献共享服务平台	北京市知春路 4 号, 100088, 62377213, www.cssn.net.cn
上海市标准化研究院·上海标准化服务信息网	上海市长乐路 1219、1227 号, 200031, 64370807, www.cnsis.info
国家科技图书文献中心·国家科技数字图书馆	理工农医界联合共建的文献保障中心, 总部在中信所, www.nstl.gov.cn
中国高等教育文献保障系统 CALIS·中国高等教育数字图书馆	国内高校联合共建的公共服务体系, 总部在北京大学, www.calis.edu.cn

1.1.3 信息检索系统

信息检索系统由计算机、数据库及相关软件(数据采集程序、索引程序、检索程序)组成。它可能只拥有一个数据库, 提供单库检索; 也可能拥有多个数据库, 既可进行单库检索, 也可进行跨库检索。根据系统数据库所含信息的性质, 可把检索系统划分为数据信息检索系统、文献信息检索系统、网页信息检索系统(网页搜索引擎), 或兼而有之。例如, 百度搜索引擎除网页搜索外, 还可提供百度百科、百度知道等数据检索服务; 谷歌的学术搜索则是典型的文献检索服务; 而有的文献检索系统中, 又包括了电子版参考工具书和企业信息数据库。

国内外常见的文献信息检索系统有:

- 维普期刊资源整合服务平台, 本书以下简称: 维普。
- 万方数据知识服务平台, 本书以下简称: 万方。
- CNKI 中国知识资源总库, 本书以下简称: CNKI。
- Engineering Village, 本书以下简称: EI。
- ScienceDirect, 本书以下简称: SD。





1.1.4 信息检索数据库

数据库是信息检索系统在数据采集并有序存储后所生成的产物。一般有数据型数据库、文献型数据库和网页型数据库等几类。

1. 数据型数据库

数据型数据库的结构属于关系型数据库,整个数据库由若干条记录(Record)组成,每条记录则划分成若干个字段(Field),记录和字段之间保持类似于表格的二维关系。表1-2为数据型数据库实例(某产品信息数据库)的片段示意图,表中行为记录,列为字段,列首为字段名,“Rec.”列为记录号字段,按先来后到顺序自动生成,其余各列中的字段内容(字段值)通过手工或计算机扫描等方法依次录入。

表1-2 某产品信息数据库的片段示意图

Rec.	Products	Company	Information	Pictures
34210	Radar sensors-VEG-APULS 61	VEGA Grieshaber KG	VEGAPULS 61 is suitable for applications in liquids in smaller vessels under easy process conditions.	
34211	Radar sensors-VEG-APULS 67	VEGA Grieshaber KG	Radar sensor for continuous level measurement of bulk solids.	
34212	Radar sensor-VEGA-PULS 63	VEGA Grieshaber KG	Radar sensor for continuous level measurement.	
34213	FMCW MICROWAVE RADAR LEVEL SENSOR MWS-20RF-CH RANGE FINDER	Wire Automatic Device Co Ltd (WADE-CO LTD.)	State of the art, high performance FM-CW Microwave Radar developed for difficult, high temperature applications in Iron and Steel making... features Rotary Microwaves and advanced mWave Software package.	

2. 文献型数据库

(1) 文献型数据库的特征

文献型数据库(简称文献数据库)的结构与数据型数据库的结构相似,也属于关系型数据库,只不过文献的格式比较固定,因此其字段组成比较规范。文献数据库的字段相当于图书馆目录中的著录项,又叫做元数据(即关于数据的数据),例如,图1-3所示的某篇期刊文献,其元数据即为该文首页上用来描述文献数据特征的题名、作者、作者单位、摘要、关键词、中图法分类号、文献标识码、DOI编号以及文献出处等信息数据。

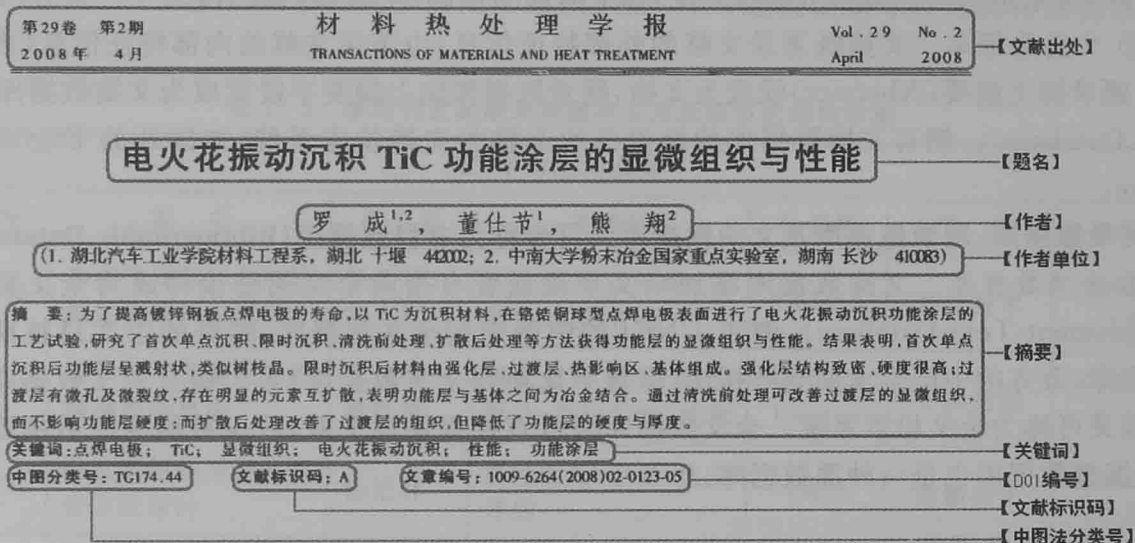


图 1-3 某期刊文献文首片段及其元数据



小贴士

什么是文献标识码?

文献标识码用来标识文献的性质,其含义如下:

- A:理论与应用研究学术论文(包括综述性论文);
- B:实用性技术成果报告;
- C:业务指导与技术管理性文章;
- D:一般动态信息;
- E:文件、资料。

什么是 DOI 编号?

DOI是“Digital Object Identifier”的缩写形式,也译为文章编号,可用来揭示数字对象的位置,例如期刊文献的DOI格式就由ISSN统一刊号、(年)、期、起始页码、页数等组成,如图1-3所示的1009-6264(2008)02-0123-05。

(2) 文献型数据库的类型

根据文献的收录对象和字段组成,文献数据库一般可分为:

① 目录数据库 目录(catalog)是对整本文献(如图书)外部特征信息的著录,目录数据库拥有记录号、书名、作者、出版地、出版社、出版年等字段,拥有目录数据库的检索系统又称为目录检索系统,例如各校图书馆的联机公共查询目录(Online Public Access Catalog,简称OPAC)。

② 题录数据库(索引数据库) 题录是对析出文献(如期刊中的个别文章、图书中的个别章节)外部特征信息的著录,题录数据库拥有记录号、题名(标题、篇名)、作者(著者、责任人、专利权人)、作者单位(机构)、关键词、分类号、源文献出处(如期刊的刊名及刊期、书名及章节名称)等字段。题录数据库的主要用途是向用户指引析出文献的出处,实际上就是一种用来查找析出文献出处的索引(篇目索引),故国外习称之为Index Database(索引数据库)。拥有索引

数据库的检索系统又称为索引检索系统,如上海图书馆的《全国报刊索引》。

③ 文摘数据库 文摘既著录文献的外部特征信息,也著录文献的内部特征信息(内容摘要)。题录加上摘要(Abstract)就成为文摘,题录数据库加上摘要字段就成为文摘数据库(Abstract Database)。拥有文摘数据库的检索系统又称为文摘检索系统,如国外的 Engineering Village。

目录数据库、题录数据库及文摘数据库又可合称为书目数据库(Bibliographic Database)。

④ 全文数据库 文摘数据库添加全文字段或者有指向全文的链接即成为全文数据库(Complement Text Database),例如,CNKI的中国期刊全文数据库、维普的中文科技期刊全文数据库、万方的中国学术期刊数据库(原数字化期刊全文数据库)等。拥有全文数据库的检索系统又可称为全文检索系统。全文数据库因可直接提供源文献,故也被称为源数据库。同理,数据型数据库也是一种源数据库。

3. 网页型数据库

网页型数据库的结构有些特别。搜索引擎的网页采集器在访问网页时,先将网页中的文本格式符、图片、音频、视频等非文本信息过滤掉,然后以文本备份的形式存入搜索引擎的缓存(Cache)服务器中,这种网页缓存文本备份(Cached Copy)又叫做网页快照(Snapshot of Web Page),网页数据库实际上就是网页快照的集合,它没有严格的字段结构,但可以把其超文本格式源文件中被<title>……</title>标记包括的范围视为网页的title(标题)字段,文本备份的文字内容视为text(内文)字段,分别编制索引。

网页数据库因收录了原始网页的备份(网页快照),故也可视为全文数据库,但它只是纯文本备份,与真实的网页还有一定距离。

1.1.5 数据库的索引

1. 整字段索引

在数据库中查找数据的方法视数据库的大小而异。小型数据库可利用计算机直接扫描,例如,在Excel数据清单中利用计算机查找radar sensor,可找出Radar sensor,radar sensor,radar sensors……,可见,扫描时的规则是letter by letter,即逐字母排比,不分大小写,也不用顾及词尾变化,前方一致即可。

大型数据库因扫描费时太多,都采用索引法,在索引中的搜索规则是word by word,即一个一个词条进行匹配。例如,索引中有词条radar sensors,则查找radar sensor时,就不可能命中该词条,除非另加有截词算符或系统默认截词。

文献数据库都是大型数据库,故对库中某些字段甚至全部字段都分别建立了索引。建立了索引的字段称为索引字段或可检字段,而被选中的可检字段则称为检索字段,又称为检索入口或检索途径。检索词与索引词的匹配操作就在检索字段的索引中进行。

传统的索引均为整字段索引,其特征是抽取完整的字段值或子字段值作为索引词(此工作称为标引),所谓子字段即多值字段(如关键词字段、作者字段)中以分号隔开的部分。将索引词与对应的记录号(库中地址)组合起来,就形成了索引款目。由于取的是整字段值或整子字段值,故此法只适用于字段值较短的刊名字段或子字段值较短的主题词、关键词、作者或分类号等字段,表1-3展示了关键词整字段索引的建立过程,抽取的是关键词字段中的子字段值。