



工程信息检索 教程

王知津 主编

于晓燕 周贺来 副主编



机械工业出版社
China Machine Press

工程信息检索 教程

王知津 主编

于晓燕 周贺来 副主编



机械工业出版社
China Machine Press

本书从工程技术角度,详细介绍信息检索的技术和使用方法,充分考虑了各种工程领域对信息检索的需要,在内容组织上进行了精心设计和安排,全书的体系结构体现了“整合中文,强化英文;突出综合,细化专业;加强训练,注重实战”的特点。全书分为四篇共10章,包括中文信息检索、综合数据库检索、专业数据库检索、特种文献检索、事实与数据检索、联机检索、因特网信息检索、综合检索实例、科研能力拓展训练等内容。

本书适合作为工程硕士“信息检索”课程的教材,也可供理工科本科生和研究生使用。对于从事工程技术领域的研究、设计和开发的专业人员来说,本书也是一本实用的参考书。

版权所有,侵权必究。

本书法律顾问 北京市展达律师事务所

图书在版编目(CIP)数据

工程信息检索教程/王知津主编. —北京:机械工业出版社, 2008. 12

ISBN 978-7-111-25227-6

I. 工… II. 王… III. 情报检索-教材 IV. G252.7

中国版本图书馆CIP数据核字(2008)第152612号

机械工业出版社(北京市西城区百万庄大街22号 邮政编码 100037)

责任编辑:王春华

三河市明辉印装有限公司印刷

2009年1月第1版第1次印刷

184mm×260mm·22印张

标准书号:ISBN 978-7-111-25227-6

定价:36.00元

凡购本书,如有倒页、脱页、缺页,由本社发行部调换
本社购书热线:(010) 68326294

前 言

如果在 20 年前提起“信息检索”，恐怕没有多少人听说过，因为那个时候信息检索还远离广大最终用户，而只是信息检索专业工作者的专用术语。当然，这并不意味着广大最终用户不需要信息检索，事实恰恰相反，人们在学习、工作和生活的各个领域里，每时每刻都在需求信息和利用信息，只不过绝大多数的检索操作都不是用户亲自进行的，而是由专职人员代替完成的。然而，20 年来，随着计算机技术、通信技术和网络技术的飞速发展，特别是 Internet 的触觉延伸到世界的各个角落，成为家喻户晓、人人皆知的大众工具，从而使信息检索也发生了翻天覆地的变化。今天再提起“信息检索”已经不是什么新鲜事了，变成了大多数人耳熟能详的常用术语。

在网络时代，信息数量急剧增长，大有“泛滥”、“污染”、“过剩”之势。如何从浩如烟海的信息海洋中寻找和获取自己所需要的信息，这是每个人都要经常面临的迫切任务。信息检索就是一门如何寻找和获取信息的学问和技艺。谁学会了信息检索的方法，谁就能够在信息海洋中遨游；谁拥有了信息检索的技巧，谁就掌握了能够打开人类知识宝库的钥匙。

对于科研工作者来说，信息检索尤为重要。科学技术的发展具有连续性和继承性的特点，任何一项研究成果的诞生都不可能不借鉴前人的成果，都不可能不受到已有成果的启发，即使是最伟大的科学家、发明家也不例外。通过信息检索可以了解和掌握国内外已有的成果，可以知道这个课题有没有人研究过，是成功了还是失败了，有哪些经验和教训，进而调整自己的研究方向和思路。只有这样，才能做到心中有数，稳操胜券，避免了重复劳动或者走弯路，节省了人力、物力、财力和时间，提高了科研效率和成功率。

2007 年 12 月 17 日，全国工程硕士专业学位教育指导委员会发出《关于加强工程硕士“信息检索”课程建设的通知》（指导委〔2007〕13 号），将“信息检索”课程增列为工程硕士的公共必修课，这充分体现了国家对信息素养教育的重视。工程硕士研究生具有其自身的特点，例如，他们当中的许多人有过几年工作经验，同时，一部分人在大学本科阶段已经学习过检索课，掌握了一些与检索相关的基本知识，这就与从来没有接触过检索课的学生有所不同。由此可见，为工程硕士开设的检索课在课程的性质、特点和方法等方面都应当与其他在校大学生有所不同。20 多年来，尽管已经编写出版了很多有关信息检索方面的教材，其中也不乏优秀之作，但绝大多数都是面向在校本科生的，而适合于工程硕士使用的教材并不多见。

我曾于 2003 年编写过一本《科技信息检索》，2005 年又翻译出版了《现代信息检索》，对信息检索有兴趣、也有一些经验，所以应出版社的邀请，组织编写了这本面向工程硕士的教材。我们在总结信息检索课程的教学实践和经验的基础上，对本书的大纲进行了充分讨论。特别强调了本书要突出三个特点：第一，主要针对工程技术领域，适当兼顾其他领域；第二，主要针对研究生层次，适当兼顾本科生；第三，主要针对有一定工作经验的工程硕士，适当兼顾工学硕士。为了体现这些特点，在章节的安排上，加大了实用性的篇幅，大大压缩理论性的篇幅；

并单独设立了“综合检索实例”和“科研能力拓展训练”两章；在章节的内容上，尽可能把数据库涵盖全，检索举例分布均匀，尽可能照顾到各个工程学科。

全书分为四篇共10章，第一篇为基础篇（第1章），用尽可能少的文字简单介绍信息检索的初步知识；第二篇为中文篇（第2章），全面、系统地介绍中文信息检索，涵盖综合性和专业性数据库以及各种文献类型、检索类型和检索方式；第三篇为英文篇（第3~8章），分章详细、系统地介绍英文信息检索，其中，第4章各节的划分参考了学科分类和专业目录，设立15个大类，尽可能全面反映工程硕士的培养领域；第四篇为提高篇（第9~10章），旨在强化学生对信息检索的综合运用能力和科研拓展能力。

本书各章节的编写者及具体分工如下：王知津（第1章）、李凌杰（第2章）、刘云（第3章）、于晓燕（4.1~4.7节）、侯延香（4.8~4.15节）、龙霞（5.1节、5.5节）、陈婧（5.2节、5.3节和5.4节）、严贝妮（第6章）、褚金涛（第7章）、谢瑶（第8章）、李维云（第9章）、周贺来（第10章）。全书书稿的初审由于晓燕和周贺来完成，王知津负责终审和定稿。

本书得以顺利完成，是与许多人的大力支持分不开的。南京大学沈固朝教授和浙江大学叶鹰教授对本书的编写大纲提出了非常有益的修改建议；清华大学韩丽风和张喜来、东北林业大学张静、天津大学刘亚茹等为本书的检索实例提供了很多帮助；机械工业出版社华章分社的编辑们对本书的编写给予了许多鼓励和指导，并提出了许多细致的修改意见，在此，我们一并表示诚挚的谢意。

在本书的编写过程中，我们参考和借鉴了大量的中外文书刊和网站资料，在此对这些参考文献作者表示诚挚的谢意。由于篇幅所限，我们未能一一列出所有参考文献，在此对未能列出的参考文献作者表示深深的歉意和诚挚的谢意。此外，我们在编写过程中，还使用了大量的中外文数据库，在此，我们谨向这些数据库商表示由衷的敬意。

虽然我们尽了自己最大的努力争取把这本教材编好，但信息检索（特别是数据库）毕竟是一个快速发展和不断更新的领域，限于编者的学识、水平和能力，缺点、疏漏和错误在所难免，恳请各位专家、学者和广大读者不吝赐教、指正，以便在本书修订时加以补充、更正和完善。

王知津

2008年8月于南开大学

教学建议

教学内容	学习要点及教学要求	课时安排
第1章 绪论	重点讲授的内容包括：信息检索的概念、原理、类型和特点；常用信息源及其特点；检索工具的类型与特点；信息检索的途径及其选择；信息检索的基本步骤。信息检索在工程技术领域中的作用与意义以及工程技术领域信息用户及其需求这两部分要结合实例讲授，使学生提高对信息检索在工程技术领域重要性的认识。信息源、检索语言、检索工具以及信息检索的一般方法可以略讲，使学生能够初步了解即可。	2
第2章 中文信息检索	主要向学生全面介绍各类型中文数据库及相关资源的检索方法，结合具体实例或课后习题讲解各数据库及相关资源的使用，通过实际操作，使学生熟练掌握中文数据库检索。该章讲授重点是中文期刊、专利、标准和学位论文数据库以及中文搜索引擎的检索与利用，其他类型的数据库及相关资源可作简单介绍。	6
第3章 综合数据库检索	该章是本课程的重点之一，所涉及的数据库可面向各工程领域的学生。讲授的重点是：Ei Compendex、Scopus、CSA、ScienceDirect、WILEY InterScience、SCI 和 SpringerLink。建议在授课过程中结合具体实例，边演示、边讲解，并安排学生上机操作。上机时，可结合课后习题进行练习，或者结合本专业选题。其他数据库可根据实际情况选讲，如 EBSCOhost、Nature、Science 等。	9
第4章 专业数据库检索	根据学生的不同专业，选择该章的相应内容来讲授。除了使用该章所涉及的数据库讲授外，还可使用本校拥有的数字资源，安排相关数据库的上机实践。例如，当授课对象为“生物医学工程”专业的学生时，除了要选择与该专业直接相关的 4.13 节中相关数据库作为授课的主要内容外，考虑到 SciFinder Scholar 数据库对该专业也是非常有用的，因此，还可以把与该专业密切相关的 4.7 节也纳入授课内容。教学采用边演示、边讲解的方法，并安排学生上机操作。	8
第5章 特种文献检索	首先应当阐述各类特种文献的特点、作用及其对科学研究的重要意义，让学生认识到特种文献检索的重要性，以提高学生的学习兴趣。重点讲授的特种文献类型包括专利、科技报告、会议文献和学位论文，对于标准文献检索可作一般性介绍。因涉及数据库较多，可根据实际情况选择部分数据库重点讲授，例如，专利文献数据库可选择德温特创新索引数据库（DII）、欧洲专利库和美国专利库来详细讲授。教学方法以实例演示为主、上机操作为辅。	8

(续)

教学内容	学习要点及教学要求	课时安排
第6章 事实与数据检索	重点讲授百科全书、手册、名录、统计资料、年鉴、书刊题录类事实与数据检索的方法,可以围绕每节所列举的数据库操作展开教学。其中,对于各数据库的概况可以简单介绍,而对于有检索示例的数据库要详细讲解,还要结合检索示例的每一个步骤,向学生演示在数据库中查找事实与数据的检索方法与技巧。该章的习题可供学生上机练习使用,也可安排学生先编写检索方案,再上机实际操作。	6
第7章 联机信息检索	重点讲授两大联机检索系统(DIALOG和STN)的使用方法。在讲授DIALOG时,要让学生全面了解DIALOG的各种服务方式、常用功能以及基本检索指令等。如果有条件,可登录到DIALOG进行演示和讲解;如果没有条件,也可选择进入Dialogweb网站,讲解Dialogweb总索引文档(DIALINDEX)的免费检索功能。STN的讲解方式与DIALOG大体相同。对于联机检索的构成、检索方式、检索步骤、发展趋势等可作为次要内容简单介绍。	4
第8章 因特网信息检索	讲授各类型因特网检索工具的特点、使用和检索方法以及性能评价等,实践性较强。教学重点是搜索引擎、网络资源目录。该章特别强调实践性教学环节,通过上机检索贴近专业的典型实例,引导学生体验不同的因特网检索工具在准确性、全面性、时效性、便捷性和响应速度等方面的差异,并保证学生完成一定数量的作业和操作练习。教学方法可以采用讲授法、讨论法和实践法等。	8
第9章 综合检索实例	该章的教学思路是通过一个检索实例,讲授课题检索的全部过程,重点讲授:检索前准备工作中的课题内容特征分析及检索策略的制定,尤其是检索词与逻辑关系的确定;检索过程中可能出现的具体问题,尤其是如何进行扩检与缩检以及如何利用数据库的辅助功能;检索结果处理中的原文获取方法以及文献阅读与分析过程的注意事项。既可以利用该章的实例进行讲解,也可以根据学生的专业情况另选一例。课后最好能够结合本校拥有的资源情况安排上机实习。	6
第10章 科研能力拓展训练	教学重点是科研项目工作流程与研究报告的撰写以及学术论文和学位论文的写作。教学难点为科技查新、成果登记、专利申请与实施。教学中不应一味地按照书本进行条理性介绍,最好采用“实例法”进行内容的讲解,以增强直观性。例如,可以以教师本人参与的课题为例,介绍科研项目流程和科研报告的写作,学术论文的撰写,以及科技查新、成果登记、专利申请和实施等。	3

目 录

前言

教学建议

第一篇 基础篇

第1章 绪论	1
1.1 信息检索的概念及意义	1
1.1.1 信息检索的概念与原理	1
1.1.2 信息检索的类型与特点	2
1.1.3 信息检索在工程技术领域中的作用与意义	3
1.1.4 工程技术领域信息用户及其需求	4
1.2 信息源	6
1.2.1 信息源的概念	6
1.2.2 信息源的类型与特点	6
1.2.3 常用信息源及其特点	7
1.3 检索语言	9
1.3.1 检索语言的概念与功能	9
1.3.2 检索语言的类型与特点	9
1.4 检索工具	10
1.4.1 检索工具的概念与功能	10
1.4.2 检索工具的类型与特点	10
1.4.3 检索工具的选择	11
1.5 信息检索的途径、方法与步骤	11
1.5.1 信息检索的途径及其选择	11
1.5.2 信息检索的一般方法	12
1.5.3 信息检索的基本步骤	12
习题	13

第二篇 中文篇

第2章 中文信息检索	15
2.1 常用中文期刊数据库检索	15
2.1.1 中文期刊全文数据库	15

2.1.2 中文科技期刊数据库	20
2.1.3 万方数字化期刊全文数据库	24
2.1.4 中文期刊数据库其他资源	26
2.2 中国专利文献检索	27
2.2.1 中国专利文献概述	27
2.2.2 国家知识产权局专利检索系统	28
2.2.3 中国专利信息中心	29
2.2.4 中国专利数据库	30
2.2.5 中国专利技术数据库	30
2.2.6 中国专利文献其他相关资源	31
2.3 中国标准文献检索	32
2.3.1 标准文献概述	32
2.3.2 中外标准全文数据库	32
2.3.3 中国标准数据库	33
2.3.4 中国标准文献其他相关资源	33
2.4 中国学位论文检索	34
2.4.1 中国学位论文概述	34
2.4.2 中国学位论文文摘/全文数据库	34
2.4.3 中国博士/优秀硕士学位论文全文数据库	34
2.4.4 CALIS 高校学位论文库	35
2.4.5 中国学位论文其他相关资源	35
2.5 中文其他特种文献检索	36
2.5.1 中文科技报告	36
2.5.2 中国政府出版物	37
2.5.3 中国会议文献	37
习题	40

第三篇 英文篇

第3章 综合数据库检索	41
3.1 Ei Compindex 数据库	41
3.1.1 概述	41
3.1.2 检索方式	41
3.1.3 检索结果浏览和处理	44

3.1.4 个性化功能	45	4.2.2 SPE 数据库	85
3.2 Scopus 数据库	47	4.3 冶金与材料工程类	86
3.2.1 概述	47	4.3.1 Maney 数据库	86
3.2.2 检索方式	47	4.3.2 其他相关资源	90
3.2.3 检索结果浏览和处理	48	4.4 机械与仪器仪表工程类	91
3.2.4 浏览来源期刊索引	49	4.4.1 ASME 数据库	91
3.2.5 个性化功能	49	4.4.2 SAE 数据库	94
3.3 CSA 数据库	51	4.5 动力与核能工程类	98
3.3.1 概述	51	4.5.1 ANS 数据库	98
3.3.2 检索方式	51	4.5.2 其他相关资源	100
3.3.3 检索结果浏览和处理	52	4.6 电气电子与信息工程类	101
3.3.4 个性化功能	53	4.6.1 IEL 数据库	101
3.4 ScienceDirect 数据库	54	4.6.2 INSPEC 数据库	106
3.4.1 概述	54	4.7 化学工程类	109
3.4.2 使用方法	54	4.7.1 SciFinder Scholar 数据库	109
3.4.3 检索结果浏览和处理	56	4.7.2 ACS 数据库	115
3.4.4 个性化功能	57	4.8 轻工纺织与食品工程类	118
3.5 WILEY InterScience 数据库	59	4.8.1 TTI 数据库	118
3.5.1 概述	59	4.8.2 FSTA 数据库	118
3.5.2 使用方法	59	4.8.3 其他相关资源	122
3.5.3 检索结果浏览和处理	61	4.9 土木建筑与水利工程类	122
3.5.4 个性化功能	61	4.9.1 ASCE 数据库	122
3.6 SCI 数据库	62	4.9.2 IWA 数据库	126
3.6.1 概述	62	4.9.3 其他相关资源	131
3.6.2 检索方式	63	4.10 交通运输与海洋工程类	131
3.6.3 检索结果浏览和处理	64	4.10.1 TRIS 数据库	131
3.6.4 最新信息通报服务	66	4.10.2 其他相关资源	135
3.7 SpringerLink 数据库	67	4.11 航空航天与兵器工程类	136
3.7.1 概述	67	4.11.1 Jane's 数据库	136
3.7.2 使用方法	67	4.11.2 其他相关资源	139
3.7.3 检索结果浏览和处理	68	4.12 环境工程类	140
3.8 其他数据库	68	4.12.1 ESA 数据库	140
3.8.1 EBSCOhost 数据库	68	4.12.2 其他相关资源	144
3.8.2 Nature 数据库	69	4.13 生物医学与制药工程类	145
3.8.3 Science 数据库	69	4.13.1 Medline 数据库	145
习题	69	4.13.2 BIOSIS Previews 数据库	145
第4章 专业数据库检索	70	4.13.3 其他相关资源	148
4.1 数理科学类	71	4.14 农业与林业工程类	149
4.1.1 SPIE 数据库	71	4.14.1 ASABE 数据库	149
4.1.2 AIP/APS 数据库	74	4.14.2 AGRIS 数据库	151
4.1.3 其他相关资源	79	4.14.3 其他相关资源	151
4.2 地质与矿业工程类	80	4.15 管理工程类	152
4.2.1 GSW 数据库	80	4.15.1 Emerald 数据库	152

4.15.2 ABI/INFORM 数据库	155	6.4 统计资料数据库	244
习题	161	6.4.1 FAOSTAT 数据库	244
第5章 特种文献检索	162	6.4.2 EIU CountryData 数据库	246
5.1 专利文献检索	162	6.5 年鉴	250
5.1.1 国外专利文献概述	162	6.5.1 The World Factbook	250
5.1.2 专利文献数据库检索	163	6.5.2 Infoplease Almanac	251
5.1.3 国际及区域组织专利文献检索	169	习题	254
5.1.4 主要国家专利文献检索	175	第7章 联机信息检索	255
5.2 科技报告检索	182	7.1 联机信息检索概述	255
5.2.1 国外科技报告概述	182	7.1.1 联机信息检索系统的构成	255
5.2.2 NTIS 数据库	182	7.1.2 联机信息检索系统的新发展	255
5.2.3 STINET 数据库	185	7.1.3 联机信息检索的步骤	256
5.2.4 NTRS 数据库	189	7.2 DIALOG 系统检索	257
5.2.5 DOE information Bridge 数据库	191	7.2.1 DIALOG 系统概况	257
5.3 会议文献检索	194	7.2.2 DIALOG 系统的常用检索指令	263
5.3.1 国外会议文献概述	194	7.2.3 DIALOG 系统的常用运算符	265
5.3.2 ISTP 数据库	194	7.3 STN 系统检索	268
5.4 学位论文检索	199	7.3.1 STN 系统概况	268
5.4.1 国外学位论文概述	199	7.3.2 STN 系统的常用检索指令	270
5.4.2 PQDT 数据库	200	7.3.3 STN 系统的常用运算符	272
5.4.3 NDLTD 数据库	204	习题	274
5.5 标准文献检索	207	第8章 因特网信息检索	275
5.5.1 国外标准文献概述	207	8.1 因特网信息检索概述	275
5.5.2 标准文献数据库检索	208	8.1.1 因特网信息检索的基本知识	275
5.5.3 国际及区域组织标准文献检索	210	8.1.2 因特网信息检索的特点	275
5.5.4 专业学会/协会标准文献检索	216	8.1.3 因特网信息检索的方法	276
5.5.5 主要国家标准文献检索	219	8.2 搜索引擎	277
习题	224	8.2.1 搜索引擎概述	277
第6章 事实与数据检索	225	8.2.2 常用综合性搜索引擎的应用	281
6.1 百科全书	225	8.2.3 常用学术性搜索引擎的应用	286
6.1.1 Encyclopedia Britannica Online	225	8.3 网络资源目录	291
6.1.2 Microsoft Encarta	227	8.3.1 网络资源目录概述	291
6.1.3 The Merck Index	228	8.3.2 常用网络资源目录检索	292
6.2 手册数据库	231	8.3.3 学科信息门户	294
6.2.1 Knovel 电子工具书	231	习题	296
6.2.2 CRCnetBASE 系列数据库	234		
6.2.3 Beilstein/Gmelin CrossFire 数据库	235		
6.3 名录数据库	240		
6.3.1 Gale-Biography Resource Center 数据库	240		
6.3.2 ProQuest Hoover's Company Records 数据库	242		
		第四篇 提高篇	
		第9章 综合检索实例	297
		9.1 课题实例与检索设计	297
		9.1.1 课题学术背景	297
		9.1.2 检索目的	297
		9.1.3 检索设计	298

9.2 课题分析	299	10.2 研究课题申报	316
9.2.1 课题内容特征分析	299	10.2.1 研究课题申报的一般程序	316
9.2.2 信息源调研与选择	299	10.2.2 科研选题的原则与依据	318
9.2.3 检索策略的制定	300	10.2.3 研究课题的类型与获取渠道	318
9.3 检索过程与结果	301	10.2.4 课题申请书的撰写	320
9.3.1 中文信息检索	301	10.3 研究报告与论文撰写	322
9.3.2 综合数据库检索	304	10.3.1 课题研究报告的撰写	322
9.3.3 专业数据库检索	307	10.3.2 学术论文的撰写	323
9.3.4 特种文献检索	307	10.3.3 学位论文的撰写	330
9.3.5 事实与数据检索	308	10.4 科技查新与成果登记	332
9.3.6 因特网信息检索	308	10.4.1 科技查新的委托	332
9.4 结果处理与分析	309	10.4.2 科技成果的鉴定	333
9.4.1 检索结果的管理	309	10.4.3 科技成果的登记	334
9.4.2 原始文献的获取	309	10.5 专利申请与专利实施	335
9.4.3 文献阅读与分析	311	10.5.1 科技成果申请专利的意义	335
习题	311	10.5.2 专利申请前的准备工作	336
第10章 科研能力拓展训练	312	10.5.3 我国专利申请的步骤	337
10.1 科研过程与科研能力	312	10.5.4 专利实施的方式与方法	338
10.1.1 科学研究的一般过程	312	习题	339
10.1.2 科研能力的概念及内容	313	参考文献	340
10.1.3 科研工作的学术规范	313		

第一篇 基础篇

第1章 绪论

本章提示

本章简单介绍信息检索的基本概念和基本理论,重点是信息源、检索工具以及信息检索的途径、方法与步骤。通过本章的学习,学生应当初步了解信息检索的主要概念、原理、类型、特点,认识到信息检索在工程技术领域中的作用与意义,理解工程技术领域信息用户及其需求,初步掌握常用信息源及其特点、检索工具的类型、特点与选择以及信息检索的途径选择、一般方法和基本步骤。

1.1 信息检索的概念及意义

1.1.1 信息检索的概念与原理

20世纪90年代前,知道“信息检索”这个术语的人还不多。随着因特网的形成、发展和普及,信息检索才被越来越多的人所知所用。

就信息检索这个概念而言,不同的使用者有不同的理解和解释,大体上可以分为两类。

第一类是广义的,对于专门从事信息检索及其系统的研究、开发和设计的少数人来说,“信息检索”可以用英文 Information Retrieval 来表达,其完整的含义是“信息存储与检索”,即 Information Storage and Retrieval。也就是说,所谓信息检索,是指将有用信息按照一定的方式组织存放起来,当用户需要这些信息时,再把它们从存放的地方查找和提取出来。因此,广义的信息检索包含存储和检索两个方面,缺一不可。

第二类是狭义的,对于广大信息用户来说,在大多数情况下,“信息检索”可以用英文 Information Searching 来表达,其准确含义是“信息查找”或“信息搜索”。也就是说,所谓信息检索,是指按照一定的方式从现有的信息集合或数据库中找到并提取所需要的信息。可见,狭义的信息检索仅指检索这一个阶段或过程。

除了专门从事信息检索及其系统的研究、开发和设计的少数人以外,其他绝大多数人使用信息检索这个术语都应该是狭义的。这就是说,包括工程技术人员在内的所有从事各行各业工作的研究、生产、服务、管理、教学等的各类人员甚至在校大学生都属于这个范围,即最广泛意义上的信息用户。

广义信息检索的基本原理如图1-1所示。在存储过程中,专门负责信息检索系统和数据库建立的人从各种各样的信息资源中,搜集有用信息,对有用信息进行主

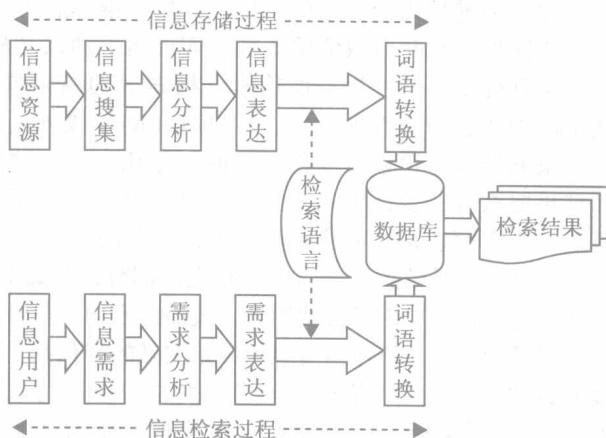


图 1-1 广义信息检索基本原理

题内容的分析,找出能够全面、准确表达主题内容的概念,借助于检索语言(通常是检索词表)把分析出来的概念转换成检索系统所采用的词语,或者直接使用自然语言,再按照一定的规则和方式将这些有用信息组织成可供检索用的数据库,并存储在一定的介质上。

检索是存储的相似过程。信息用户在工作、学习和生活中产生了信息需求,为了检索并获取自己所需要的信息,用户必须对自己的需求进行主题内容的分析,找出能够全面、准确表达主题内容的概念,也要借助于检索语言(通常是检索词表)把分析出来的概念转换成检索系统所采用的词语,或者直接使用自然语言,再按照一定的检索规则和方式,制定检索策略,构造检索式,从数据库中查找并获取自己所需要的信息,最后,输出检索结果。当然,检索的全过程还应当包括对检索结果进行评价、反馈,或许还可以重新制定检索策略,重新构造检索式,反复进行检索,直至检索出满意的结果为止。

狭义信息检索的基本原理就是广义信息检索中信息检索过程的那部分原理。

信息检索基本原理中的核心部分是用户所使用的检索词或者由检索词和运算符所组成的检索式与数据库中的检索词及其逻辑关系之间的比较和匹配机理。两者相匹配的信息被检索出来(命中),不相匹配的信息被拒绝。

1.1.2 信息检索的类型与特点

1. 按照检索对象划分

(1) 文献检索

文献检索的对象是文献,例如,检索“太阳能电池”方面的文献。这里所说的“文献”是指文献单元,即包含一个完整内容的单元,如一篇论文、一本图书、一份报告等,而忽略其物理载体(如纸介质、磁介质、光介质)、出版形式(如图书、期刊、报纸)、加工深度(如一次文献、二次文献、三次文献)等。进一步讲,这里所说的“文献”可以是完整的原始文献,也可以是原始文献的替代品,如一条目录款目、一条文摘款目或一条索引款目。归根到底,文献检索的目标是检索出原始文献或原始文献的替代品。

(2) 数值检索

数值检索的对象是数值,数值的范围不限于数值本身,还包括图形、图表、数学表达式、化合物分子式、化学结构式等非数字型的数值。数值检索的目标是检索出可以直接使用的数值,例如,检索“尼罗河的长度是多少”,“喜马拉雅山的高度是多少”。数值检索有时也叫数据检索。

(3) 事实检索

事实检索的对象是事实,例如,检索“长江哪一年汛期的水位最高”,“克隆羊最早是由谁研制成功的”。与文献检索和数值检索不同,事实检索一般不能通过简单检索直接提供问题的答案,而必须进行比较复杂对比、分析、推理后才能得出最终结果。

在国外,有时对数值检索和事实检索不加以区分,把两者概括在数据检索或事实检索之下。

文献检索是一种相关性检索,检索出的结果具有不确定性。也就是说,检索结果出来后,还不能确定它们是否满足要求,只有在阅读或浏览了原始文献之后才能确定。因此,只能说检索出来的结果与检索课题是相关的。相比之下,数值检索和事实检索是确定性检索,检索出来的结果要么有,要么无;要么是,要么否,直接回答用户的具体问题,毫不含糊。此外,如上所述,事实检索是3种检索类型中最复杂的。

文献检索是信息检索的核心和主体部分,与数值检索和事实检索相比,内容更丰富,方法更灵活,是信息用户最经常使用的。按照内容的完整性,文献检索又可以分为书目检索和全文检索。所谓书目检索,是指检索对象为原始文献的替代品,即文献线索,而不是原始文献本身,要想阅读原始文献,还必须依据文献线索去进一步找到和获取原始文献。书目检索通常借助于文摘数据库、索引数据库、目录数据库来完成。所谓全文检索,是指检索对象为原始文献本身,检索结果就是原始文献,可以直接阅读和使用原始文献。全文检索通常借助于全文数据库来完成。

2. 按照计算机检索技术划分

与联机检索相对应的是脱机检索,这是计算机检索的最早技术,作为计算机检索的一个发展阶段,脱机检索在计算机检索历史上仍然占有一席之地,但目前已经很少使用了。

(1) 联机检索

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电话线路等数据通讯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。由于联机检索系统的功能较强、数据库的质量较好,所以联机检索的费用较高。联机检索的鼎盛时期是20世纪60年代中期到80年代中期,至今仍被使用。

(2) 光盘检索

光盘检索分单机系统和联机系统两种,光盘单机检索系统自成系统,提供单个用户使用,通常由微机、光盘驱动器、光盘数据库及相应的检索软件和驱动软件组成。光盘联机检索系统是在光盘网络的环境下运行的,光盘网络受到光盘塔和局域网的支撑,在局域网内提供多个用户使用,由服务器管理。光盘数据库大多由联机检索提供商提供,因此两者的检索方法大体相同。光盘检索费用低,但数据更新慢。光盘检索的鼎盛时期是20世纪80年代中期到90年代初,至今仍被使用。

(3) 网络检索

基于搜索引擎技术的网络检索是随着Internet的兴起和普及而出现的。Internet上的信息非常广泛、丰富,但又非常杂乱、无序,网络信息资源分布在世界各地的主机上,信息量巨大,动态更新,主要依靠搜索引擎获取。Web搜索引擎是当今网络检索工具的主流,不仅能够提供文本检索,还可以提供图形、图像、音频、视频、动画等多媒体检索。目前,网络信息检索已经成为信息检索的主要途径。

1.1.3 信息检索在工程技术领域中的作用与意义

“科学技术是第一生产力”。在知识经济兴起和经济全球化趋势迅速发展的今天,要把我国建设成伟大的社会主义强国,最主要的就是要坚持创新。国家创新体系包括3个方面:知识创新、技术创新和制度创新。知识创新体系的主要功能是把人类对自然界进行的基础研究和应用基础研究的成果扩散和传播到需要它们的技术创新体系中去,启发并产生新产品、新工艺、新材料领域;技术创新体系的主要功能是创造、革新、学习和传播新技术,开发研究高新技术,促进科学技术向现实生产力转化,实现综合国力和国际竞争力的增强;制度创新体系的主要功能是进行体制改革,制定新的政策,保证知识创新体系和技术创新体系的正常运转,并对各类创新资源进行合理配置,推动经济持续发展。其中,知识创新是基础,技术创新是目的,制度创新是纽带。

技术创新主要体现在工程领域,虽然工程领域也包括一部分知识创新和制度创新。从技术角度讲,工程领域同生产直接相关,主要从事产品设计开发、工艺制造、维护、业务管理等活动。工程领域的主体是工程技术人员,工程技术人员主要分布在两大部门:一个是研究、设计部门,承担较复杂项目的研究、设计工作,或者承担重要的研究课题或主持和组织重大工程项目设计,解决本专业领域的比较复杂的或关键性的技术问题;另一个是生产、技术管理部门,运用现代生产管理和技术管理方法,解决在生产过程或综合技术管理中本专业领域重要技术问题。为了完成好上述工作,工程技术人员必须及时了解和熟悉本专业领域国内外发展现状和趋势,及时掌握和采用国内外先进技术,提高研究、设计、生产、技术管理水平和经济效益,而要实现这些目标都与信息检索技能的培养和提高密切相关。

概括起来,信息检索在工程领域中的作用主要体现在以下几个方面。

1. 为提高科研效率提供有效途径

众所周知,任何科研工作都离不开4种资源:人力资源(科研队伍)、物力资源(材料设

备)、财力资源(资金经费)和信息资源(情报资料), 四者缺一不可。科学技术的发展具有连续性和继承性的特点, 任何一项研究成果的诞生都不可能不借鉴前人的成果, 都不可能不受到已有成果的启发, 即使是最伟大的科学家、发明家也不例外。通过信息检索可以了解和掌握国内外已有的成果, 可以知道这个项目有没有人搞过, 是成功了还是失败了, 有哪些经验和教训, 进而调整自己的研究方向和思路。只有这样才能做到心中有数, 稳操胜券, 避免了重复劳动或者走弯路, 节省了人力、物力、财力和时间, 提高了科研效率和成功率。

2. 为科研立项和课题研究提供科学依据

科研立项不仅要避免重复, 还要保证新颖性。一般地讲, 任何科研工作都是从课题论证起步的, 要想把课题论证好, 必须进行信息检索。通过信息检索可以全面、准确地了解同类课题发展的来龙去脉, 哪些问题已经解决, 哪些问题正在研究, 哪些问题尚未研究, 目前的水平如何, 将来的发展趋势是什么, 等等。所有这些问题都必须通过信息检索来解决, 然后才能做好科学的充分的论证, 从而保证立项课题的高起点。课题论证阶段所进行的信息检索是科技查新的一部分。在课题研究展开过程中, 还会遇到许多事先没有估计的新问题, 解决这些问题也需要进行信息检索, 对原有课题加以补充和完善, 直至课题研究圆满成功。

3. 为科技成果的鉴定、评估、验收、转化、奖励提供客观依据

当研究课题完成后, 通常也要进行信息检索, 以便为科研管理部门或成果评审委员会对科技成果进行鉴定、评估、验收、转化、奖励等提供客观的参考依据, 从而判定科研成果的创新性、在国内外的技术水平和地位、应用价值, 证明其科学性和可靠性, 以避免对科技成果评定的失真、失准, 消除主观性。这属于另一种类型的科技查新。

4. 为自主知识产权的取得和保护提供保障

许多科技成果, 特别是工程技术领域的成果, 根据其经济价值, 往往都要申请专利, 取得知识产权保护, 这对于企业和个人都具有十分重要的经济意义。从某种意义上讲, 拥有自主知识产权的多少, 是衡量一个国家经济竞争力的标准, 也是衡量一个企业竞争力的标准。知识产权制度是促进技术创新、加速科技成果产业化、提高经济竞争力的一项重要法律制度, 许多国家和企业都把取得和保护自主知识产权提升到发展战略的高度予以重视。不管是发明专利还是实用新型专利、外观设计专利, 不管是产品发明还是方法发明, 都必须符合专利法要求的“三性”: 新颖性、先进性、实用性。因此, 在申请专利时, 在发生专利纠纷时, 都要进行信息检索, 以便取得和保护自己的专利权。这又是一种科技查新。

5. 为提高自学能力、实现终身学习铸造“利剑”

信息检索不仅是一种方法, 更是一种技能。我们所处的时代是信息爆炸的时代, 随着社会网络化进程的加速, 信息每时每刻都像潮水般涌来, 我们已经陷入了信息的汪洋大海。在全球经济加速的大环境下, 科技成果从发明到推广应用的周期大大缩短, 知识的有效性也在逐步缩短。要想跟上时代的步伐, 除了学习以外别无选择。对于科技人员来说, 他们的绝大部分知识都是在实践中学习和积累的, 如果他们能够熟练掌握和运用信息检索技能的话, 他们就能够不断地补充知识, 及时地更新知识, 调整自己的知识结构, 提高解决科技难题的能力, 适应科技和生产的发展。“授人以鱼”只能享用一时, 只有“授人以渔”才能受益终身。

1.1.4 工程技术领域信息用户及其需求

工程技术领域信息用户是指从事研究、设计、生产、技术管理的各级各类人员。这些人员虽然都具有工程技术领域信息用户的一般特征, 但由于他们所从事的具体工作和所承担的具体任务不同, 在信息需求方面也表现出不同的特征。为了叙述方便, 我们把工程技术领域信息用户划分为4种类型: 科学研究人员、设计开发人员、生产技术人员、技术管理人员。

1. 科学研究人员信息需求特点

科学研究人员包括从事自然科学基础研究、应用基础研究和技术科学研究活动的一切人员。

他们的信息需求特点主要表现为:

1) 专业性。他们的研究工作主要是基础性的,研究领域相对狭窄,但研究难度较大,短期内不容易出成果,对信息内容的要求比较专深。

2) 系统性。基础性的科学研究要求科研人员所掌握的知识必须是系统的、完整的,对信息的要求不能是分散的、零碎的。

3) 连续性。基础性研究一般不能在短期内完成,需要持续一段时间,有时甚至需要相当长的时间,在这段时间内,必须连续地研究,往往不能中断或者时断时续。对信息的时间性要求不是很高。

4) 多样性。正因为所从事的研究是基础性的,才要求信息类型和形式的多样性,除了最常用的期刊论文以外,还需要学位论文、会议论文,甚至科技报告。

2. 设计开发人员信息需求特点

设计开发人员属于技术创新人员,直接面向实际生产,从事产品和工艺的规划、设计、开发、试验、试制和发明创造活动,把基础研究的成果转化为产品和工艺,创造新技术、新方法、新产品、新材料。设计开发人员的信息需求特点主要表现为:

1) 行业性。设计开发人员的工作直接关系到企业生产的产品和企业提供的服务在本行业中的竞争力,他们对信息的要求通常限制在本行业内,一般不怎么关心本行业外的信息。

2) 广泛性。不仅需要本行业新工艺、新方法、新产品、新材料、新能源等方面的信息,也需要应用基础研究和科学技术研究成果方面的信息,还需要市场和消费者方面的信息。

3) 可靠性。设计开发人员的工作是直接面向市场、面向产品的,与科学研究人员的探索性工作不同,不允许在技术上、产品上失败,所要求的信息必须是可靠的、有把握的。

4) 实用性。他们获取信息的目的是解决生产实践中的具体问题,着眼于实际应用和经济效益,所需要信息的类型和形式是期刊、专利、标准、产品样本、手册和实物等。

5) 时间性。受到产品生命周期、技术生命周期以及生产和市场需要的紧迫性影响,设计开发人员对技术信息和市场信息的要求是最新的,过时的信息毫无价值。

3. 生产技术人员信息需求特点

生产技术人员身处生产实际的第一线,负责把开发出来的新产品投入生产并把设计出来的新工艺的实际应用技术实现,还要负责产品质量检验和标准化管理,直接与生产流程、生产工艺、设备使用和维护、产品制造打交道,生产过程中可能出现的一切问题他们都会遇到,也都需要他们迅速加以解决。由于生产技术人员与设计开发人员联系非常紧密,但又直接联系到生产第一线,所以他们的信息需求与设计开发人员有一定的相似性。不同的是,生产技术人员对信息的要求更加可靠和实用,更加具体和有针对性,因为他们要立竿见影地解决问题,容不得拖延,否则将贻误生产和商机。

4. 技术管理人员信息需求特点

技术管理人员包括各级科技主管部门、各级各类研究设计单位、各级各类技术市场以及企业技术部门等的管理人员,这些人员在不同层次、不同类型的技术管理工作中形成了各自的信息需求,其信息需求特点主要表现为:

1) 综合性。为了实现科学管理,技术管理人员除了需要本行业的科技信息以外,更多的是需要与本行业科学技术发展有关的方针政策、法律法规、自然资源、社会经济、交通运输、环境保护、文化教育、金融贸易、市场行情等的比较宏观的信息,从政府到行业,再到企业,从国内到国外,涉及面很广。

2) 决策性。技术管理人员利用信息的目的不是解决具体的科学技术问题,而是进行管理决策,所需要的信息应当是比较宏观的、经验型和解决方案型的,有助于做出科学决策。

3) 准确性。一旦管理决策失误,必将导致不可估量的损失,其后果是相当严重的。任何

不符合客观实际的主观的或误传的信息都会给管理决策造成不良影响。技术管理人员对信息的要求是准确无误。

4) 简洁性。做出正确的管理决策不需要长篇大论的冗长的论述性信息,信息只要准确、客观、简单明了即可,技术管理人员对信息形式的要求以报告、简报、消息、动态为主。

1.2 信息源

1.2.1 信息源的概念

人类对于自然界和社会的认识无不来源于社会实践。信息源就是信息的来源。人类在认识世界和改造世界的过程中所获得和产生的各种认识成果都是信息的来源,这些成果通常以一定的物质形态被记录下来。记录下来的原始信息和知识还可以被进一步加工处理成次生的信息和知识,不管信息和知识是原生的还是次生的,都可以被需要它们的人所利用。

由此说来,我们在上面所提及的工程技术领域中的科学研究人员、设计开发人员、生产技术人员、技术管理人员,他们既是信息和知识的生产者和创造者,也是信息和知识的加工者和使用者。工程技术人员在科研、生产和管理活动中源源不断地产生了大量的信息和知识,同时又每时每刻地需要、获取和利用信息。因此,可以把他们看成是广义的信息源。

然而,从信息检索的意义上说,由于检索的对象是信息,即信息的内容和形式,而信息又是可以被传递的,所以,在信息检索中,我们把承载信息的客体称为信息源。

1.2.2 信息源的类型与特点

1. 按照信息的表现形态划分

1) 口语信息源。这是指人们以口头语言的表现方式所产生和传递的信息,如谈话、讲话、发言、报告、演讲、授课、讨论、歌唱等。这类信息源没有被记录下来,存在于人脑的记忆中。它们的特点是直接、及时、针对性强并有可能现场反馈和互动,缺点是稍纵即逝,不便于记录、积累、保存和重复,容易误传。

2) 体语信息源。这是指通过人的肢体语言的表现形式所产生和传递的信息,如表情、手势、姿态、动作、舞蹈、杂技、服装表演、体育比赛等。这类信息源的特点与口语信息源相似,但直观性更强,生动活泼,极富感染力,印象深刻。

3) 实物信息源。这是指通过实际物品本身原有的物理形态所展现出来的信息,如产品、商品、样品、物品、模型、标本等,既可以是人工制品,也可以是天然制品。实物信息源主要是通过展览、展销、展示、陈列以及反求工程来获得的。这类信息源的特点是真实、可靠、易分析、易模仿,而且直观性最强。

4) 文献信息源。这是指通过文字、符号、图形、图像、音频、视频、代码等表现形式记录在各种物质载体上的信息,由此而形成的记录统称为文献。其形态多种多样,如手抄本、印刷本、缩微品、录音带、录像带、磁带、磁盘、光盘以及网络介质等。应当指出,文献的概念要从广义上理解,不限于纸制品。这类信息源的特点是数量巨大,应用广泛,历史悠久,便于保存、积累和反复使用,构成了信息源的主体,也是信息检索的主要对象。

2. 按照信息的记录手段划分

1) 手写信息源。这是指采用手抄和刻写的记录手段的最古老的信息源,如古代的甲骨文、金石文、帛文、竹木文等,现代的如手稿、日记、笔记、书信、会议记录、设计草图、原始档案等。这类信息源的特点是原始性和稀缺性。

2) 印刷信息源。这是指以纸张为载体、以印刷或手写为记录手段而产生的传统形式的信息,也叫书本型信息源,如图书、期刊、报纸等。这类信息源的特点是便于阅读和流通,符合人们获取信息的习惯,但存储密度小、体积大,存放和管理占用空间和人力。

3) 缩微信息源。这是指以缩微平片或胶片等感光材料为载体、利用光学技术为记录手段而产生的信息形式。这类信息源的特点是存储密度大、体积小、寿命长,便于存放和管理,但借