

科技信息检索

主 编 周和玉 郭玉强

副主编 张 倩 周运凤

陕西人民教育出版社

科技信息检索

主 编 周和玉 郭玉强
副主编 张 倩 周运凤

编 者 王冠炬 刘红昆 朱远坡
杜莹琦 张 倩 郑 琳 周运凤
周和玉 岳 红 郭玉强 谢万明

陕西人民教育出版社

内容简介

本书共分 19 章,全面系统地介绍了科技信息基本知识,信息检索的基本原理与方法,中外文检索工具的使用,以及会议、报告、专利等特种文献的检索、计算机信息检索、光盘检索和网络信息检索等内容。为了加强大学生的素质教育,本书重点放在培养学生解决实际问题和利用检索工具和掌握现代的检索手段的技能上。本书可作为大学生、研究生及函授生的教材,也可供工程技术、科教、图书情报人员参考。

主题词:科技信息——信息检索——检索方法

(陕)新登字 004 号

图书在版编目(CIP)数据

科技信息检索/周和玉,郭玉强主编. —西安:陕西人民教育出版社
ISBN7-5419-7081-6

I. 科…

Ⅱ. 周…

Ⅲ. 科技信息—检索

IV. G760

科技信息检索

周和玉、郭玉强主编

出版发行:陕西人民教育出版社(西安长安路南段 376 号)

经 销:新华书店

开本:787×1092 毫米 1/16

版次:1999 年 8 月第 1 版

印刷:陕西高陵印刷厂

印张:12.5 字数:280 千字

印次:1999 年 8 月第 1 次印刷

印数:1—5000

ISBN 7-5419-7081-6/G·6244

定价:15.00 元(包括实习指导书)

序

随着科学技术的飞速发展,科学技术文献的数量与类型也在急剧增加。凡从事现代教育、科学研究以及生产开发的工作者都不可避免地要从这些科技文献中汲取大量的科技情报,借以拓展思路,激发创造。然而在浩瀚的知识和信息的海洋里查寻特定的符合自身工作需要的知识情报,的确不是一件容易的事。只有那些通过学习,掌握了科技文献检索的理论和方法的人们才会得心应手,查寻自如。当今的科技工作者普遍地认识到,科技检索已是大学生们必不可少的专门知识之一。

科技文献检索是在四十年代末期,在传统的图书馆学、目录学的基础上与现代语言学、信息学相结合,逐步形成的一门相对独立的学科。掌握这门学科首先可以防止课题的重复研究,避免失败或重复劳动。检索工作做得越充分,选题就越准确,研究工作获得成功的可能性就越大。其次有助于提高科研效率。由于现代科学技术综合交错,彼此渗透,面对浩如烟海的文献资料,科技工作需要的文献在查找中必然受到严重干扰,耗去大量的时间和精力。而开设文献检索课的目的之一,就是要有效地缩短人们搜索情报的时间,进而缩短科研的周期,达到多出成果快出成果的目的。再者科技文献检索课又是一门传授获取新知识的方法、提高获取新知识的技能、实践性很强的课程,掌握这门课程有利于人们随时补充、更新知识、改善不合理知识结构,提高解决科技难题的应变能力。因此有人说这是一门一次受教育,终身受益的培养开拓型人材不可缺少的课程。

经过专家们的多年努力,把文献检索课列入教育计划已成为现实。多年的教学实践证明,它对于提高教学质量,拓宽学生的知识面,提高科研能力起到了不可估量的作用。这次由武汉交通科技大学图书馆编写出版的《科技信息检索》教材对培养新一代交通运输人才必将起到更大的作用。

田维成

1999年6月

前 言

我们所处的时代,是一个科学技术迅猛发展的时代,是一个信息知识激增的时代,信息已成为人类社会赖以生存和发展的三大支柱之一,它已渗透到社会的各个角落。因而,也可以说,获取信息,利用知识已成为人们生存的本领之一。作为科技工作者、大学生更应掌握信息获取的基本知识和基本技能,以便更快更全面地获取当今世界的最新科技信息,掌握新知识,不断完善和发展自己,为社会作出更大贡献。

本书是按原国家教委高等教育司“《文献检索课教学基本要求》教高司1992[44]号”文件要求编写的。全书共十九章,除信息检索的基本理论和基本检索技能及主要中外文检索工具介绍等内容外,还增加了计算机信息检索,光盘检索和网络信息检索等内容介绍。本书内容新颖,资料翔实,叙述详略得当,有较强的实用性。既有利于培养读者的实际操作能力,又能了解信息检索的最新发展。

本书是由长期从事信息检索课教学和信息咨询工作的同志编写的。其中,王冠炬编写第十、十一章,朱远坡编写第十六、十七章,刘红昆编写第八、十五章,张倩编写第三、四章,杜莹琦编写第十九章,周运凤编写第七、九章,岳红编写第六、十四章,郑琳编写第十八章,郭玉强编写第一、二章,谢万明编写第十三章,周和玉编写第五、十二章。全书由周和玉、郭玉强、张倩、周运凤统稿,最后由周和玉、郭玉强整理、定稿。在编写和出版过程中,武汉交通科技大学图书馆原老领导龚其福、刘新华两位老师以及总支书记田维成副研究员、副馆长童国强副研究馆员、教材中心负责同志给予了多方面的指导和支持,在此一并致谢。特别感谢杜云波同志在本书出版过程中给予的大力支持。

由于时间仓促,加之编者水平有限,书中如有疏漏和不妥之处,请专家和读者给予指正。

编者

1999.6.30

目 录

序

前 言

第一章 绪 论	(1)
第一节 信息知识、情报、文献	(1)
第二节 学习、信息检索的意义和作用	(2)
第三节 信息检索基本概念	(4)
第四节 信息检索效果的评价	(7)
第五节 科技文献的级别、特点和类型	(8)
第二章 检索语言及其发展趋势	(14)
第一节 检索语言	(14)
第二节 检索语言的发展趋势——分类主题一体化	(18)
第三章 检索的步骤与方法	(22)
第一节 分析研究课题	(22)
第二节 选择检索工具	(23)
第三节 确定检索途径	(25)
第四节 选择检索方法	(26)
第五节 查找文献线索	(28)
第六节 索取原始文献	(31)
第四章 中文检索工具	(32)
第一节 概 述	(32)
第二节 主要中文检索工具	(32)
第三节 中文检索工具的编排结构	(35)
第四节 我国主要检索工具选介	(38)
第五节 检索方法	(39)
第五章 美国《工程索引》简介及使用方法	(42)
第一节 概 述	(42)
第二节 《工程索引》词表	(45)
第三节 检索途径与实例	(47)
第六章 英国《科学文摘》及使用	(50)
第一节 概 述	(50)
第二节 《科学文摘》的结构编排与著录格式	(50)
第三节 《INSPEC》叙词表	(55)
第四节 检索途径与检索步骤	(56)
第七章 《英国海运技术文摘》及使用	(60)
第一节 概 述	(60)

第二节 《BMTA》的结构体系	(60)
第三节 《BMTA》的编排与著录格式	(61)
第四节 《BMTA》检索途径与检索实例	(63)
第八章 美英《金属文摘》及使用	(64)
第一节 概 述	(64)
第二节 《金属文摘》结构编排及著录格式	(64)
第三节 《金属文摘索引》结构编排及著录格式	(67)
第四节 《金属文摘年度索引》结构编排及著录格式	(68)
第五节 《合金索引》(Alloys Index)结构编排及著录格式	(68)
第六节 《ASM》冶金词表	(69)
第七节 《金属文摘》检索途径与实例	(70)
第九章 美国《科学引文索引》及使用	(73)
第一节 概 述	(73)
第二节 结构编排及著录格式	(73)
第三节 使用方法与检索实例	(77)
第十章 科技报告及其检索	(79)
第一节 概 述	(79)
第二节 PB 和 AD 报告及其检索工具	(80)
第三节 《GRA&I》的检索方法	(84)
第四节 NASA 报告及其检索	(88)
第五节 DOE 报告及其检索	(90)
第十一章 会议文献及其检索	(92)
第一节 概 述	(92)
第二节 学术会议检索工具	(93)
第三节 检索途径及检索实例	(97)
第十二章 专利基础知识	(100)
第一节 专利制度和专利发明	(100)
第二节 专利文献	(104)
第三节 专利文献的分类	(104)
第十三章 中国专利文献信息检索	(107)
第一节 中国专利制度	(107)
第二节 中国专利文献类型	(107)
第三节 中国专利文献的结构编排	(110)
第四节 中国专利文献检索途径	(112)
第十四章 英国《世界专利索引》及使用	(114)
第一节 概 述	(114)
第二节 WPI 出版物体系	(115)
第三节 《世界专利索引》编排体例	(116)

第四节 WPI 的一般检索途径和检索实例	(120)
第十五章 其它专利文献检索系统	(125)
第一节 各国专利公报	(125)
第二节 INPADOC 检索系统	(127)
第三节 美国《化学文摘》	(130)
第四节 《世界多国发明》	(131)
第十六章 计算机情报检索基础知识	(132)
第一节 概 述	(132)
第二节 计算机情报检索的基本原理及功能	(133)
第十七章 国际联机情报检索	(138)
第一节 国际联机情报检索系统简介	(138)
第二节 联机检索命令	(141)
第三节 联机检索步骤及实例	(144)
第十八章 光盘检索系统	(150)
第一节 光盘技术简介	(150)
第二节 中文科技期刊数据库光盘检索系统	(152)
第三节 Ei 光盘检索系统	(157)
第十九章 INTERNET 网络信息检索技术	(165)
第一节 概 述	(165)
第二节 网上信息检索技术	(170)
第三节 网络信息资源	(180)
参考文献	(191)

第一章 绪 论

第一节 信息、知识、情报、文献

我们所处的时代,是一个高新科技迅猛发展,在前所未有的层次上不断产生、传递和利用信息情报的时代。信息已成为现代社会赖以生存和发展的三大支柱(信息、物质、能源)之一。因此,有人称我们的时代为信息时代,那么,究竟什么是信息、知识,什么是情报,它们与本课程大量讨论的文献有着什么关系,这是我们首先需要了解的问题。

一、信息

信息是一种十分广泛的概念,它在自然界、人类社会以及人类思维活动中普遍存在。不同事物有着不同的特征。这些特征通过一定的物质形式(如声波、电磁波、图像等)给人带来某种信息。例如,人的大脑通过感觉器官所接收到的有关外界及其变化的消息,就是一种信息。因此,信息可以定义为:生物以及具有自动控制功能的系统,通过感觉器官和相应的设备与外界 进行交换的一切内容。

二、知识

知识是人类社会实践经验和认识的总结,是人的主观世界对于客观世界的概括和如实反映。知识是人类通过信息对自然界、人类社会以及思维方式与运动规律的认识,是人的大脑通过思维加工重新组合的系统化信息的集合。因此,人类不仅要通过信息感知世界、认识和改造世界,而且要将部分所获得的信息升华为知识。可见,知识是信息的一部分。

三、情报

关于情报的定义,国内外学术界众说纷纭,至今还没有定论的说法。但大家的基本共识为:情报是指传递着有特定效用的知识。知识性、传递性和效用性是情报的三个基本属性。

1. 情报的知识性

人们在生产和生活活动中,通过各种媒介手段(书刊、广播、会议、参观等),随时都在接收、传递和利用大量的感性和理性知识。这些知识中就包含着人们所需要的情报。情报的本质是知识,可以说,没有一定的知识内容,就不能成为情报。

2. 情报的传递性

情报的传递性是说知识要变成情报,还必须经过运动。钱学森说情报是激活的知识,也是指情报的传递性。人们的脑海中或任何文献上无论贮存或记载着多少丰富的知识,如果不进行传递交流,人们无法知道其是否存在,就不能成为情报。情报的传递性表明情报必须借助一定的物质形式才能传递和利用。这种物质形式可以是声波、电波、印刷物或其它,其中最主要的是以印刷物等形式出现的文献。

3. 情报的效用性

运动着的知识也不都是情报,只有那些能满足特定要求的运动的知识才可称之为情报。例如,每天通过广播传递的大量信息,是典型的运动的知识。但对大多数人来说,这些广播内容只是消息,而只有少数人利用广播的内容增加了知识或解决了问题,这部分人可将其称之为情报。

四、文献

文献是用文字、图形、符号、声频、视频等技术手段记录人类知识的一种载体。因此,根据不同载体形式和读取方式,文献大体可分为印刷型、缩微型、机读型和视听型等四种主要形式。

由上述可见,知识是信息中的一部分,情报是知识中的一部分,文献是知识的一种载体。文献不仅是情报传递的主要物质形式,也是吸收利用情报的主要手段。

信息、知识、情报、文献四者关系如图 1-1 所示

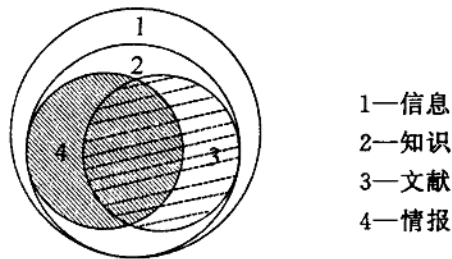


图 1-1 信息、知识、情报、文献的关系

第二节 学习信息检索的意义和作用

一、信息检索的意义

在大学教育中,对大学生来说,最主要的是五种能力的培养,即自学能力、研究能力、思维能力、表达能力和组织管理能力。信息检索是培养学生能力的基本技能和方法之一。

对于硕士生来说,在学习阶段中,主要是学会如何独立从事研究工作。在查找文献的基础上,善于总结前人经验,在前人成就的基础上有所创新。

对于一个博士来说,主要是通过信息检索,培养在总结前人经验的基础上选择具有创造性的研究方向的能力,能够开辟新的研究领域。

对于一个科技工作者来说,在开始研究课题之前,首先要进行大量的调研工作,以便了解和继承前人的成果,摸清该学科或课题的现有水平和发展趋势,以及与其相关学科之间的关系。

然而,当代科技文献呈现的状态使得科技人员检索适用的信息变得愈来愈困难,我们处于一个知识、信息量激增的年代。国际教育发展委员会主席埃德加·富尔说过:“我们也不能刻苦地、一劳永逸地获取知识了。而需要终身学习如何去建立一个不断演进的知识体系——学会生存。”教会学生建立一个不断演进的知识体系,培养学生的适应能力和独立科研能力,正是现在高等教育中存在的薄弱环节,亟需加强。在对学生进行传授基本知

识的同时,要着重培养其独立吸收和运用信息的能力。

要掌握具有吸收信息能力的知识,必须克服三个方面的语言障碍,即自然语言障碍、学科语言障碍和图书信息系统中的检索语言障碍,主要表现在不懂信息检索语言,不掌握信息检索的方法和技能,以致不能有效地利用图书信息系统地去检索有关科技文献资料,以调整知识结构,解决实际问题。如果学生仅仅具备一定的外语和学科专业知识,而没有一定的信息检索与利用的能力,就会面对汪洋大海般的科技文献资料陷入找不到、读不完的困境。因此,在高校中开设“科技信息检索”课,有目的地培养学生的情报意识和信息吸收能力,使其具备自我知识更新和获取最新信息的能力,其重要性就在于此。

二、信息检索的作用

1. 有利于减少课题的重复研究、提高科研成功率

在开始着手研究一项课题之前,必须掌握这个课题是如何提出来的,前人曾在这方面做过什么工作,如何做的,有何成果和经验、教训,还存在什么问题,以及相邻学科的发展对研究这个课题提供了哪些新的有利条件等等。也就是说,必须通过检索来掌握有关科技信息,进而摸清国内外的发展水平与动态,并以此作为研究的起点,借鉴前人成功的经验和教训,做到有充分材料根据论证课题的可行性,正确地制定研究方案。因而,检索工作做得越充分,选题就越准确,研究工作获得成功的可能性就越大。所以,检索工作是课题研究不可缺少的先期工作,也是防止重复研究、免遭失败的必要手段。例如:日本高能物理研究所由于借鉴了外国的经验及其它因素,它的第一台高能加速器投资比国外同类装置节省1/2,人力省3/4。相反,欧洲共同体在技术开发中,由于信息不灵,忽视了他人已有的成果,使15%的研究工作要么重复,要么不能受益。我国的信息工作一直比较落后,所以重复研究的现象比较严重。一方面重复研究国外已有的技术;另一方面,国内各机构之间相互重复研究及引进、生产的现象也很严重。

2. 有助于节约时间、提高科研效率

随着科学技术的不断发展,科学技术文献的数量与类型也在急剧增加。据资料统计,目前全世界大约有80000种科技期刊,每年发表的约600万篇论文。全世界每年出版专利说明书(包括等同专利)已达100多万件,每年出版会议录达1万种以上。面对这浩如烟海的文献,科技工作者需要的文献很难查到,而不需要的文献却严重干扰查寻者的视线,耗去了他们大量的精力和时间。因此不少科学家惊呼这是一种“情报危机”、“情报污染”和“情报爆炸”。同时,由于现代科学纵横交错,彼此渗透,使得文献的专业性质也不十分固定,也给科技工作者掌握与利用专业文献增加了难度。

美国对科研人员工作时间的调查表明:美国科技人员在一项研究工作的全过程中,用于计划思考的时间占7.7%;用于试验研究的时间占32.1%;用于数据处理的时间占9.3%;而用于搜集情报资料和发表成果的时间占50.9%;由此可见,搜集资料发表成果的时间占去研究时间的1/2。如果能将这部分时间减少到最低限度,使科研人员把主要精力和时间用于构思和研究,那就等于增加了科研人员,或延长了科研人员寿命。这是发展科学技术的一个巨大潜力。我们开设文献检索课的目的之一,就是要有效地缩短科技人员搜集资料的时间,进而缩短科研周期,达到多出成果,快出成果的目的。

在当今世界,提高科研效率,加快科研速度的意义还在于使相同科研课题在国内外竟

争中处于有利位置。专利法规定,对相同的发明成果,按“先申请原则”授予专利权。即多项相同发明,只授予第一个先申请人专利权,其后申请的发明作为已知技术处理。显然,如果忽视科研速度,即使科研获得了成功,但由于发明失去了时间上的“新颖性”,也会变成无效劳动,给国家带来损失。

3. 有利于培养复合性、开拓性人才

迈向 21 世纪的现代科学技术的发展日新月异,每日每时都会有所发现、有所发明、有所创造、有所前进。随着时间的推移,不完善的方法被比较完善的方法所代替。现在,科研成果从发明到推广应用的周期大大缩短,知识的有效期也在逐步缩短。因而科技信息也出现了知识“失效”的问题。据美国工程教育协会估测:美国受过高等教育的科技人员所具有的科技知识 12.5%是在大学阶段获得的;87.5%是在工作岗位上学习积累的。估测还指出:如果大学毕业后五年之内不学习补充新知识,原有的知识 50%将陈旧失效;十年之内不学习补充新知识,原有的知识将 100%陈旧失效(即由先进的知识变成一般知识)。这就是说一个成熟的科技人员,他的知识绝大部分是在实践中学习积累的。如果不继续学习,不更新知识,原有的知识将会失效。另外,大学生上课用的教材,从编写到出版,再到学生手中,要经过 3—5 年,甚至长达十年时间,书本上的知识是静止的,并有些已经老化了,因此应将学生从静态知识引向动态知识,扩充、深化课堂知识,从书本知识中解救出来,并引导学生主动地去跟踪收集国内外的科学技术的新成果,进行研究探索,培养科学研究的能力。正因为如此,我国高等院校强调培养和提高学生的自学能力和分析研究问题的能力。开设信息检索课程的目的,就在于培养学生的上述能力,为这提供一把打开知识宝库的“钥匙”,主动地去觅取更深更广的新知识。信息检索知识,作为一种方法和技能,还有利于人们适应科技和生产发展随时补充、更新知识,改善不合理的知识结构,提高解决科技难题的应变能力。有人说这是一门一次教育终身受益的课程。

总之,信息检索知识和技能,已成为人们知识结构中不可缺少的最重要的组成部分。把信息检索课程列入教育计划,努力学习信息检索知识和操作技能,对于提高教学质量,培养复合型、开拓型人才具有十分重要的意义。

第三节 信息检索基本概念

一、信息检索的定义

信息检索(Information Retrieval)——是指将信息按一定的方式组织和贮存起来,并根据信息用户的需要找出有关的信息的过程。所以,它的全称又叫信息存贮与检索(Information Storage and Retrieval)。这是广义的信息检索;狭义的信息检索则仅指该过程的后半部分,即根据课题的需要,主要借助于检索工具,从信息集合中找出所需信息的过程,相当于人们所说的信息查寻(Information Search)。

作为检索对象的信息,它有不同的形式,有的以文献形式出现,有的以数据和事实出现。根据检索对象形式的不同,信息检索又分为文献检索、事实检索和数据检索。

1. 文献检索(Document Retrieval):是以文献(包括文摘、题录和全文)为检索对象的一种检索。即凡是查找某一主题、某一时代、某一地区、某一著者、某一文种的有关文献,以

及回答这些文献的出处和收藏处所等等,均属于文献检索范畴。例如:设计快速双体客船有哪些参考文献?就属于某一主题的文献检索。文献检索是一种相关性的检索,检索的结果是文献线索,还必须再进一步找到这些文献,阅读这些文献,然后才能筛选出所需要的情报。

2. 数据检索(Data Retrieval):这是以数据为对象的一种检索。即查寻包括在文献中的某一数据、公式、图表以及某一物质的化学分子式等等。例如:查某种型号柴油机的最大功率是多少,就属于数据检索。科学数据有两种形式:一是数值型数据,如实验数据与工业技术数据;二是非数值型的数据,如化合物分子式、化学结构式、工业产品名称及规格等等。数据检索是一种确定性的检索,检索时,要么有,要么无,要么是正确,要么是错误的。检索的结果是科技工作者直接可以利用的情报。

3. 事实检索(Fact Retrieval):这是以事实为对象的一种检索。即查寻某一事物发生的时间、地点及过程等等。例如:世界上最长的隧道是哪条?该隧道何时建设?它位于什么地方等等。这些提问均属事实检索。事实检索与数据检索一样,是一种确定性检索,检索的结果是可以供科技人员直接利用的情报,而不是一种文献线索。

以上三种检索类型的主要区别在于:数据检索与事实检索,都是要检索出包含在文献中的信息本身,而文献检索,则是要检出包含所需信息的文献。

二、文献检索的基本原理

文献检索是由文献的存贮与检索的两个不可分割的过程所组成(如图 1—2 所示)。

1. 文献存贮过程:把具有一定价值的文献,在主题分析的基础上,经过标引处理(即给文献注检索标识),按一定的规定和方法把它们系统地排列起来,并记入或输入检索工具中去。

2. 检索过程:根据文献用户的检索需要,对检索提问,进行主题分析,并将分析的主题内容进行标引处理(即给提问要求注检索标识),然后,使用检索工具,按照一定的检索方法查找所需文献线索。

综上所述,文献检索概念的完整意义应该包括存贮和检索两个方面的含义。因为没有文献的存贮,就没有文献的检索。存贮是为了检索,检索必须存贮。

为了确保文献检索的质量,在文献存贮或检索之前,必须把文献或提问的主题(中心)找出来,并形成文献或提问主题概念。标引是将上述主题概念转换成检索语言(即用分类号或主题词代替主题概念),按照一定的规则和方法进行分析著录的全过程。

主题分析的深度与广度,将直接影响标引的网罗性和主题词的专指性。标引的网罗性是指衡量标引时认识文献或提问的主题内容的广度而言。如果主题分析愈深透,那么抽出的主题概念愈多,标引的网罗性愈高,检索时,相关主题的文献就能检索出来,因而查全率较高;主题词的专指性是指其准确表达主题的能力,具有两个以上的概念的复合主题而言。选用专指性强的主题词愈多,则检索出来的文献针对性愈强,查准率愈高。查全率(检出相关文献量与存贮相关文献量之比)愈高,说明漏检相关文献少。但过高的查全率检出不不对口的文献就会增多,因而查准率(检出相关文献量与检出文献总数之比)就相应的降低,它说明误检的文献增加。反之亦然。对于检索来说,漏检是影响检索质量的主要因素,误检会降低检索的效率。总之,标引时,要根据存贮或检索的要求,适当控制标引的网罗性

和主题词的专指性,以提高文献检索质量。

文献检索的基本原理就是检索者将检索提问的标识与存贮在检索工具中的文献特征标识进行比较。结果凡是文献特征标识与检索提问标识相一致,或者文献特征标识包含了检索提问标识,那么,具有这些特征标识的文献就从检索工具中输出。输出的文献线索就与检索者所需的文献线索大致吻合。

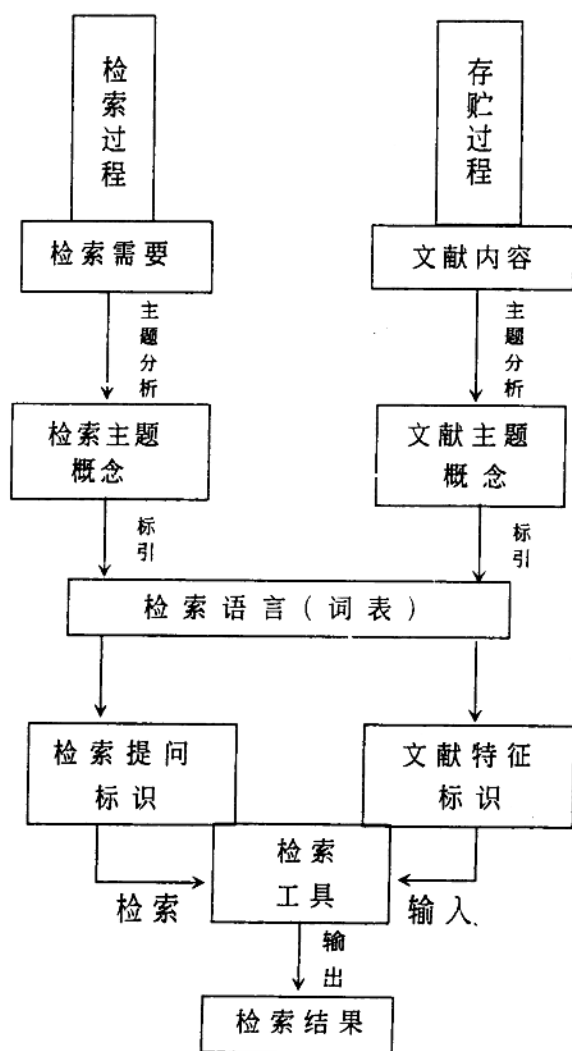


图 1—2 文献检索的基本过程

第四节 信息检索效果的评价

信息检索效果,就是利用检索系统进行检索服务时所获得的有效结果。它对用户利用信息的效果产生直接的影响,是对检索系统的性能和质量检验的尺度。评价信息检索效果,目的是为了准确掌握检索系统的各种性能水平,分析影响检索效果的因素,调节检索策略,改进检索系统的性能,提高检索效果,满足用户信息检索的需求。

一、评价指标

评价指标是衡量检索系统性能和检索效果的标准,一般包括检全率、检准率、漏检率、误检率等四项指标。

1. 检全率。它是指检出的相关文献量与检索系统中相关文献总量的比率,是衡量信息检索系统检出相关文献能力的尺度。可用下式表示:

$$\text{检全率} = \frac{\text{检出相关文献量}}{\text{系统中相关文献总量}} \times 100\%$$

2. 检准率。它是指检出的相关文献量与检出文献总量的比率,是衡量信息检索系统精确度的尺度。可用下式表示:

$$\text{检准率} = \frac{\text{检出相关文献量}}{\text{检出文献总量}} \times 100\%$$

评价信息检索系统的检索效率主要指标是检全率和检准率,即检索系统中的全部相关文献都被检出,检出的文献全部是相关文献。然而,由于许多因素的影响,在实际检索中,检全率和检准率是不可能达到100%的,而是存在着一种互逆关系,即在同一检索系统中提高检全率,检准率则会降低,反之,检准率提高,检全率则会下降。

3. 漏检率。它是指漏检相关文献量与检索系统中相关文献总量的比率,是衡量信息检索系统漏检文献总量的比率,是衡量信息检索系统漏检文献程度的尺度。可用下式表示:

$$\text{漏检率} = \frac{\text{漏检相关文献量}}{\text{系统中相关文献总量}} \times 100\%$$

4. 误检率。它是指误检(检出不相关)文献量与检出文献总量的比率,是衡量信息检索系统误检文献和程度的尺度。可用下式表示:

$$\text{误检率} = \frac{\text{误检文献量}}{\text{检出文献总量}} \times 100\%$$

评价信息检索系统的检索误差主要指标是漏检率和误检率,误差越大,效率越低,检索系统的性能就越低;误差越小,效率越高,检索系统的性能就越高。由此可见,产生漏检和误检的原因是影响信息检索系统效果的主要因素。

影响检索效果的主要原因

产生漏检和误检的原因是多方面的,如检索语言、检索系统的编制质量,标引深度和检索策略等等都将直接或间接地影响检索效果。

1. 产生漏检的主要原因:

(1) 检索语言不能全面地显示文献主题概念之间的等级和相关关系,没有排除检索语言中的同义现象,造成同一主题文献分散而致漏检;

(2)检索系统收录文献不全,检索途径不齐全;

(3)文献的标引深度不够,前后不一致或遗漏了原文的重要概念或选词不当等;

(4)检索人员没有具备选择正确检索词以表达文献主题的能力以及把选出的检索词逻辑地完整地组合在一起的能力等等。都是造成误检的原因,是影响检准率的主要原因。

三、提高检索效果的措施

提高检索效果,必须克服漏检和误检产生的原因,保证检全率和检准率,其主要措施有:

1. 提高检索语言质量。检索语言是运用检索系统进行检索的工具,它的质量高低,对检索质量有重要的影响。特别是检索语言的专指度,直接影响标引和检索文献的准确度。因此,加强对检索语言的控制,健全概念组配规则,完善词表的结构,充分揭示语义关系,是提高检索效果的重要措施。例如,加强词语规范,明确词语含义和范围,消除一词多义、同义词分散、同音或同形不同义的现象;严格组配方法的采用,合理地使用标识符号;编制必要的范畴表、词族表和轮排表,准确地揭示词语之间的关系。

2. 提高检索系统编排质量。检索系统是信息检索的物质基础,根据学科和专业的需要,广泛全面及时地收录各种相关文献信息,著录内容详尽准确,辅助索引齐全完备,系统结构科学合理,可提供多种检索途径,方便、迅速、准确、全面是一个质量较高的检索系统应具备的质量标准,也是提高检全率和检准率的重要措施。

3. 提高文献标引质量。文献标引是检索工作的基础和准备阶段,是建立检索系统的必备条件,文献的标引质量直接影响着检索系统的编制质量,是决定检索效果的重要因素,在标引工作中,能否全面反映和正确揭示文献主题,直接影响到检全率和检准率。因此,加强文献标引的深度和准确性,保持其系统性和一致性,是提高检索效果的必要措施。

4. 提高检索策略制定水平。检索策略是一种技巧性很强的检索方案,不同的检索方案得到的检索效果有着很大的差别。制订科学、全面、合理、细致的检索策略,必须具备较高的选择和运用检索语言、检索系统的能力和水平,选择和运用检索途径和检索方法的技巧,以及在检索过程中能及时调整、修改检索策略的能力,做到正确揭示文献主题,全面反映检索主题。因此,要制定一个好的检索方案,不仅要求检索人中具备较高的专业素质,也要求用户具有一定的文献检索常识,能全面、正确地表达检索要求,使检索人员能准确地制定最佳检索策略,这样才能有效地降低漏检率和误检率,提高检索效果。

第五节 科技文献的级别、特点和类型

一、科技文献的级别

根据文献的内容和其加工级次与加工方式,可将科技文献划分为一、二、三次三个级别。

1. 一次文献

也称一级文献、系指原始文献。这是著作者在科学研究、教学和生产实践中以成果为依据而创作撰写的文献。诸如专著、报刊论文、研究报告、会议文献、学位论文、研究报告、会议文献、学位论文、专利说明书、技术档案、技术标准、科技报告等,多属一次文献。只要

是原始的著述,无论是何种文献类型或载体形式,都为一次文献。文献的内容是确定文献级别的依据。一次文献具有较高的参考和使用价值。

2. 二次文献

也称二级文献。在一次文献的基础上,按一定的原则进行系统的检索、组配、加工、整理、报导和组织所形成的文献,称为二次文献。如书目、索引、文摘、题录等,均属二次文献。由于这类文献具有检索和报道一次文献的功能,故又称之为检索性文献和报道性文献。它能系统地反映一次文献信息,为读者提供检索所需文献的线索,是查找利用一次文献的检索工具。

3. 三次文献

也称三级文献。它是指在利用二次文献的基础上,对一次文献进行筛选,再将所选用的有关文献的内容进行综合、分析、研究和评述而编写出来的文献。它又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、大全等。三次文献源于一次文献,又高于一次文献,是一种再创性文献。

近几年来,有人提议将文献划分为四级,除一、二、三次文献外,另需再加零次文献。所谓零次文献,是指文献创造者与使用者直接对话,在非文献系统中进行传递的信息,简言之,即通过非正式过程交流的文献。各国学者对文献结构层次的划分标准和方法的看法不尽一致。我们将科技文献划分为3个结构层次,也只能是一种近似的划分。就科技文献的性质和用途而言,这应是一种较为客观的科学的划分方法。

二、科技文献的特点

科学技术的进步,促进了科技文献的发展。现代科技文献的发展,具有以下明显的特点:

1. 数量庞大,增长迅速

科学技术的蓬勃发展,科研成果的大量涌现,科学交流的频繁开展,极大地丰富了科技文献的宝库。据不完全统计,全世界每年出版的图书有80万种以上;科技期刊近8万种,发表的期刊论文在600万篇以上;提出的专利说明书100万种左右;国际会议文献1万种以上;技术标准逾20万件。科技文献数量之多,浩若烟海。全世界各种文献量之和将超过1.2亿种,平均每天出版文献达30万件,其中科技文献占有相当大的比重。

科技文献的出版发行,不仅数量庞大,而且增长的速度也令人惊叹。据统计,非科技内容的文献每30—50年增长1倍,而科技文献平均每7—8年就要增长1倍,某些尖端科学领域和新兴学科的文献,其增长的幅度更大,增长的速度更快。如原子能科学、环境科学和计算机科学的文献,每2—3年就翻一番。

2. 形式复杂,文种多样。

随着科学的发展,声、光、电、磁等现代技术和化学塑胶新材料的广泛应用,现代出版物发生了历史性的变化。出现了纸书印刷品与缩微资料、声像资料、机读资料、光盘资料等多种文献载体并存的趋势,从而结束了纸印刷品一统天下的局面。尽管非纸文献载体具有生产快、体积小、质量轻、信息密度大、保存时间长等诸多优点。但是,由于纸型文献载体具有生产、加工、整理、保存、传递简单的优点、且读取方便,不受严格的限制。可以预料,在今后一个比较长的时间内,纸书印刷型科技文献仍将占主导地位。多种文献载体相互补充、