



教育部人才培养模式改革和开放教育试点教材
高等医学院校卫生事业管理专业教材

Yixue Wenxian

Jiansuo

医学文献检索

■ 主编：谢志耘

北京大学医学出版社

教育部人才培养模式改革和开放教育试点教材
高等医学院校卫生事业管理专业教材

医学文献检索

主编：谢志耘

编者：陈奇敏 高 琴 李春英
沈 霞 杨丽莉

北京大学医学出版社

YIXUE WENXIAN JIANSUO

图书在版编目 (CIP) 数据

医学文献检索/谢志耘主编. —北京: 北京医科大学出版社, 2002.8 (2007.11 重印)
教育部人才培养模式改革和开放教育试点教材. 高等
医学院校卫生事业管理专业教材
ISBN 978-7-81071-239-2

I. 医… II. 谢… III. 情报检索 - 高等学校 - 教材 IV. G252.7

中国版本图书馆 CIP 数据核字 (2002) 第 042881 号

医学文献检索

主 编: 谢志耘

出版发行: 北京大学医学出版社 (电话: 010-82802230)

地 址: (100083) 北京市海淀区学院路 38 号 北京大学医学部院内

网 址: <http://www.pumpress.com.cn>

E - mail: booksale@bjmu.edu.cn

印 刷: 莱芜市圣龙印务书刊有限责任公司

经 销: 新华书店

责任编辑: 赵 蔚 责任校对: 焦 娟 责任印制: 张京生

开 本: 787mm × 1092mm 1/16 印张: 13.75 字数: 350 千字

版 次: 2002 年 8 月第 1 版 2007 年 11 月第 4 次印刷 印数: 20001 - 23000 册

书 号: ISBN 978-7-81071-239-2

定 价: 18.80 元

版权所有, 违者必究

(凡属质量问题请与本社发行部联系退换)

前 言

文献检索课是培养学生的信息意识,掌握用手工方式和计算机方式从文献中获取知识和情报的一门科学方法课,是提高学生自学能力和独立研究问题能力的工具课。本课程的任务是使学生了解各自专业及相关专业文献的基本知识,学会常用的手工文献检索工具和计算机文献检索工具的使用方法,懂得如何获得与利用文献情报,增强自学能力和研究能力。

十多年来,文献检索课已经成为高校一门独立的课程,正式列入各类专业、各层次学生的教学计划,深受学生们的欢迎和重视;也受到了教师和科研人员的关注和青睐;是否具有独立获取与利用各种信息的能力,已成为跨世纪、具有创新能力人才的重要标志之一。

随着现代通讯技术和计算机技术的飞速发展,以及网络技术和 Internet 的广泛普及与应用,查找文献信息的工具和方法发生了巨大的变化,获取文献信息的手段日益方便和简化,传统的利用印刷型检索期刊进行手工检索的方式逐渐被计算机文献信息检索方式所取代,文献检索课教学的重点也向计算机文献检索转移。因此,编写结构合理、内容丰富实用,并适合于当前教学需要的文献课教材是目前的当务之急。

本教材是为卫生事业管理专业的学生专门编写的。针对该专业的特点和教学要求,对教材内容、结构进行了规划和设计,强调理论联系实际,具有深入浅出、文字简明等特点。为适应信息电子化、网络化时代的要求,在内容上压缩手工检索比例,突出基本知识和方法,适当扩充计算机检索比例,尤其增加了当前流行的光盘数据库、网络数据库和 Internet 信息检索与利用方面的知识。从实用的角度出发,还重点讨论了美国《科学引文索引》的检索方法以及特种文献的利用,对于非重点的手工检索工具刊只作简单介绍,使之更适合于卫生事业管理专业的教学需要。另外,对文献信息的分析和利用、学术论文和综述的写作也作了详细介绍。

本教材在编写过程中得到了北京大学医学图书馆尹源研究馆员、陈荔子研究馆员以及解放军医学图书馆陈界教授的指导和大力帮助,特致谢意。

由于编者学识和水平所限,加之时间仓促,书中可能有不少错讹或不妥之处,祈望各位同行和专家批评指正。

编 者

2002 年 1 月

目 录

第一章 概论	(1)	第二节 计算机检索的基本原理	(112)
第一节 基本概念	(1)	第三节 文献信息数据库的基本概念	(113)
第二节 文献的基本概念	(3)	第四节 计算机检索方式	(115)
第三节 文献检索基本知识	(8)	第五节 计算机检索步骤与方法	(118)
练习题	(18)	练习题	(123)
第二章 国内文献检索刊物	(20)	第七章 光盘数据库检索	(124)
第一节 国内检索刊物概况	(20)	第一节 光盘及光盘检索	(124)
第二节 《中文科技资料目录》 (医药卫生)	(25)	第二节 MEDLINE 光盘检索	(127)
第三节 《中国医学文摘》	(30)	第三节 HealthSTAR 光盘检索	(138)
练习题	(33)	第四节 CBMDISC 光盘检索	(140)
第三章 国外文献检索刊物	(36)	练习题	(154)
第一节 国外检索刊物概况	(36)	第八章 Internet 信息检索与利用	(156)
第二节 美国《医学索引》	(39)	第一节 Internet 概述	(156)
练习题	(57)	第二节 Internet 的主要服务功能	(160)
第四章 特种文献检索	(59)	第三节 网上信息资源的类型与特点	(174)
第一节 专利文献及其检索	(59)	第四节 网上信息资源指南及搜索引擎	(176)
第二节 标准文献及其检索	(73)	第五节 网络数据库检索	(181)
第三节 科技报告及其检索	(77)	练习题	(189)
第四节 学位论文及其检索	(81)	第九章 文献信息的调研	(190)
第五节 会议文献及其检索	(84)	第一节 概述	(190)
练习题	(90)	第二节 基本方法与步骤	(192)
第五章 美国《科学引文索引》	(92)	第三节 文献信息分析成果的表述	(200)
第一节 《科学引文索引》概况	(92)	练习题	(207)
第二节 《科学引文索引》的结构编排与著录格式	(94)	附录 MeSH 副主题词使用范围	(208)
第三节 检索途径及应用举例	(102)	参考文献	(214)
练习题	(108)		
第六章 计算机检索概论	(110)		
第一节 计算机检索发展史	(110)		

第一章 概 论

〔教学内容〕

- 一、基本概念
- 二、文献的基本概念
- 三、文献检索基本知识

〔教学要求〕

- 一、描述信息、知识和情报的基本概念
- 二、简述文献的定义和属性
- 三、陈述科技文献的类型
- 四、阐述科技文献的特点
- 五、解释文献检索的概念
- 六、知道文献检索的目的和意义
- 七、描述文献检索原理
- 八、解释检索语言的概念
- 九、比较不同类型的检索语言
- 十、解释检索工具的概念
- 十一、列举检索工具的基本类型及一般结构
- 十二、熟悉文献检索的基本方法和途径
- 十三、熟悉文献检索的基本步骤

随着全球知识经济的迅猛发展，人类正逐渐进入一个信息时代。为了适应信息时代的要求，我们每个人都应该不断丰富知识内容，拓展多种学科领域。在浩瀚的知识海洋中，如何迅速全面地了解我们所要掌握的知识？掌握知识信息的检索技术与方法是非常有必要的。文献检索课是一门培养学生自学能力、提高研究素质的科学方法课或工具课。通过本课程的学习使学生了解各自专业及相关专业文献的基本知识，学会常用的手工检索工具、计算机检索工具及参考工具书的使用方法，掌握获取原始文献的主要方法及初步整理文献资料的方法，懂得如何获得与利用文献情报，增强自学能力和研究能力。

第一节 基本概念

在学习文献检索课的基本原理和检索方法之前，首先介绍文献检索的一些基本概念：信息、知识、情报以及三者的相互关系。

一、信息

在我们所处的世界中，信息无处不在，无时不有。比如去股票市场可以了解股市信息；

到人才市场可以了解人才需求信息；到电器商行可以了解电器信息；到软件市场可以了解软件信息；浏览经济导报可以了解经济信息等。信息不仅包括人与人之间的信号的交换，而且还包括人与自然界之间，人与机器之间，机器与机器之间的信号的交换，以及动物界与植物界的信号交换。由一个机体传递给另一个机体，甚至由一个细胞传递给另一个细胞，由一个分子传递给另一个分子的任何运动方式，也都被视为信息的传递。信息几乎存在于一切领域中。在当今经济、科技迅速发展的时代，人们已经处在信息的包围之中。如果不及时用信息来充实自己，就要被信息社会淘汰。

随着信息时代的到来,更加丰富了信息学的理论。信息的内容逐渐增多,信息的形式可以图、文、声三种并存,信息的传递速度有了迅猛的发展。人们不仅从宏观的角度来研究信息,而且也从微观的角度来研究信息,大到天体,小到分子水平来研究自然界信息的规律和特性。

究竟什么是信息?众说纷纭,不同学者从不同角度对信息做出各种定义,难以有统一的定论。哲学家认为信息是认识世界的依据;数学家认为信息是一种概率;物理学家说信息是“熵”;而通信家则把信息看做是“不定度”的描述。信息论的创始人香农(C.E.Shannon)认为:信息是组织程度,它能使系统的有序性增强,减少破坏、混乱和噪声。

信息是无形的,它是客观事物的运动状态和特征的反映,是人们认识事物发展的基础。每件事物不同的运动状态和特征会产生不同的信息反映,包括社会信息、生物信息、技术信息等。在医学上,各种体征的出现和变化是机体生理和病理状态的信息反映,各种化验结果的数据也是机体生理和病理状态的信息反映。事物是不断变化发展的,那么它的运动状态和特征也应随着不断变化而发展。因此,信息也是不断变化发展的。但人们发现和认识信息都受着生产力、科技水平及认识手段的限制,社会历史的发展是在不断获取信息、认识信息、传递和利用信息、创造信息的过程中,通过信息认识世界与改造世界。在医学领域中通过各种疾病的不同信息来认识千差万别的疾病。

在人类已经进入信息社会的今天,信息被视为如同能源、空气和水一样重要。信息资源被视为与自然资源同等重要的第二资源。人们将一个国家信息化程度的高低作为衡量一个国家发展水平的标志。一个国家科学技术发展水平、经济持续发展都与信息的开发和利用密切相关。

二、知识

汉语词典中,知识解释为:人们在改造世界的实践中所获得的认识和经验的总和。知识是人类对自然界、人类社会中各种现象、规律的信息反映进行思维分析,加工提炼,经过系统化、理论化的过程。也就是人的大脑通过思维重新组合的系统化的信息集合。因此,系统化、理论化的信息就称为知识。信息是知识的源泉、材料,知识是信息的理论和结果。例如某一症状、体征诊断为某一疾病,这种症状和体征是该疾病信息的反映,该疾病是症状和体征的信息升华,这种信息升华就是疾病的诊断知识。因此,信息是知识的源泉和基础,知识是信息的升华。

三、情报

情报最初产生于军事领域,主要是探察敌情的报告,称之为情报。情报是人们在一定的时间内为一定的目的而传递有使用价值的知识或信息。这种情报具有保密性、时效性、传递性的特点。因此,情报是知识的一部分,是特定的知识或信息。随着历史、社会以及科学技术的不断发展,情报的作用也不断地变化,由军事转移到科技、经济、社会服务等各领域。

现已转化为人们获取知识、信息的一种重要手段。

情报的种类有多种。按服务对象不同,可分为军事情报、科技情报、战略情报、战术情报等;按传递媒介分为文字情报、实物情报、声像情报;按传递范围分为大众情报和专门情报。

四、信息、知识、情报三者的相互关系

信息是知识的源泉;知识是系统化、理论化的信息;情报是活化的知识和信息,能为人们所利用,其涵义关系是信息>知识>情报。信息要成为情报,一般要经过选择、综合、分析和研究加工过程,即经过知识的阶段才能成为情报,因此它必须具有三种基本要素:知识、传递和效益。知识是情报的实体,传递是情报的表现形式,效益是情报的结果。情报就是信息经过综合、筛选、逻辑思维、重新组合的系统化知识。

第二节 文献的基本概念

一、文献的定义

自古以来,人类的知识有三种存储形式:一种是存储在人们的大脑中;另一种是存储在一定的实物中;还有一种是以文字、图形、符号等记录手段记录在一定的物质形态上。要想获得人脑中的知识必须直接接触,通过思想交流方能获得;要想获得实物中的知识,必须具备一定的实物,并通过对其研究,才能掌握它;存储在一定的物质形态上的知识,只要没有语言障碍即可获得。

“文献”一词,在我国有几千年的历史。随着科技、社会的发展,“文献”的概念也在变化。辞海的定义为:“原指典籍与贤者”。今专指具有历史价值的图书文物资料,亦指与某一学科有关的重要图书资料。国家标准《文献著录总则》对文献的定义为“文献是记录有知识的一切载体。”现在,一般认为:“凡属于人类的知识,用文字、图形、符号、声频、视频等手段记录保存下来,并用以交流传播的一切物质形态的载体,都统称为文献。”

从上述文献的定义可以看出,知识是文献的实质内容,载体是文献的外在形式,而记录是联系知识与载体的手段。

文献是本课程的主要研究对象。

二、科技文献的类型

科技文献的类型可从文献内容加工的深度、文献的载体、出版的形式、获取的难易程度等不同的角度和不同的标准来划分。

(一) 根据加工层次划分

1. 一次文献 即原始文献,是作者根据自己的工作和研究成果而写成的文章,也可称原始论文。不论撰写是否参考或引用他人的著作,不论载体和出版类型,均属一次文献。

一次文献特点是内容有创新性,含有前所未有的发明创造,或者一些新的见解与理论,是科学技术有所前进的标志。一次文献是对知识的第一次加工,是信息的基础,也叫信息源。

一次文献包括:期刊论文、研究报告、会议录、专利说明书、学位论文等。

2. 二次文献 是对一次文献进行收集、分析、整理并按照其外部特征或内部特征(篇

名、作者、作者地址、刊名、出版年、卷、期、页、分类号、内容摘要等)按一定的规则加以编排,供读者检索一次文献之用。

二次文献通常是由图书信息机构组织人力物力编辑出版的,它起着汇集文献、提炼文献和提供文献线索的作用。它是查找一次文献的线索和桥梁。

二次文献包括:目录、索引、文摘等。

3. 三次文献 是科技人员在利用二次文献的基础上,对一次文献阅读、分析、归纳、整理和推理,进行概括、论述,重新组织、加工提炼成文字,供人们了解某一学科或专题的进展,了解其过去、现在和预测未来的发展趋势的文献。

三次文献是在充分研究已发表文献的基础上,对某一专题已取得的成果、进展、或研究水平与动态进行总结和概括,对当前研究具有指导意义。篇末附有该专题的大量参考文献,为读者提供了该专题主要文献的线索,因此又具有文献检索工具的功能。

三次文献一般是由专家写成,水平较高,专业性强。

三次文献包括:综述、评论、述评、进展、动态、年鉴、专著、指南等。

从一次文献到二次文献、三次文献的过程是一个由博而精,由分散到集中,由无组织到系统化的过程。从文献检索来说,一次文献是检索的对象,二次文献是检索的手段与工具。

4. 零次文献 一般认为是形成一次文献之前的信息、知识、情报,即未经记录或未公开于社会的最原始的文献,或没有正式发表的文字材料,如书信、手稿、笔记、记录等。

(二) 按载体形式划分

1. 印刷型 (printed form) 文献 是指印刷到纸张上保存下来的文献。是传统的记录知识、信息的方式。其优点是便于直接阅读,可广泛流传;缺点是体积大,存储密度低,占用空间多。

2. 缩微型 (micro form) 文献 是以感光材料为载体,用摄影的方法把文献的影像体积缩小,记录在胶卷或胶片上。其优点是体积小、容量大、成本低、保存时间长,便于复制、携带。缺点是阅读不太方便,使用时必须借助专门的阅读机。目前最常用的是缩微胶卷 (microfilm) 和缩微平片 (microfiche)。一张国际标准型缩微平片以 1:24 的缩放比例可存储 98 页纸本型文献,读者可通过阅读机阅读并通过复印机放印成原始纸本读物。

美国一些图书馆从 20 世纪 70 年代起逐渐广泛应用,将过期文献进行缩微复制,既便于保密又方便存放,同时,图书馆可腾出更多的空间和书架来放新的书刊。

3. 视听型 (audio-visual form) 文献 是指记录声音和图像的文献,包括:唱片、录音带、录像带、幻灯片等。其优点能听其声、观其形,比较直观、真切,便于理解、掌握。视听型文献容易保存,可以反复使用。例如,记录有心脏杂音的录音带,记录一个手术过程的录像带等。其缺点是也必须借助于录音机、录放机、放映机、幻灯机等播放设备。

4. 电子型 (electronic form) 文献 又叫机读型文献,是随着计算机技术不断发展而发展起来的。此文献是以数字形式将信息存储在磁盘、磁带、光盘或网络等介质上,并通过计算机或远程通讯进行阅读的文献。它们具有高的信息存储密度和存取速度,并具有电子加工、出版和传递功能。电子出版物的产生开始了人类历史上快速、高效的知识生产和传播。人们通过计算机阅读、编辑、出版、检索和获取信息。电子出版业的迅猛发展将大大加速社会的信息化过程。

电子出版物的出现是社会信息化的一个里程碑,它有广阔的发展前景。它的产生并不意味着对其他一切信息载体的完全取代,各种载体的产生和存在有其特定的环境、条件和需

要。因此,各种载体将在相当长时间内共存,相互补充、渗透,发挥各自的优势,共同促进信息的繁荣与人类的文明。

(三) 按文献出版类型划分

1. 科技图书 (book) 大多数是对已发表的科研成果、生产技术的概况和总结。一般是经过著者对原始材料加以选择、鉴别和综合之后写成的。其内容比较成熟、系统、全面、可靠。如果想获得某一学科全面、系统的知识或对陌生的学科进行初步的了解,阅读图书是一个非常有效的途径。

科技图书是一种重要的读物,它能提供综合性科学知识,在教育和培养科技人才方面作用显著。但是,科技图书的出版周期比较长,信息传递较慢,因此,在科技人员使用信息中占的比例不是很大,它们一般不是最主要的检索对象。科技图书一般分为两种类型。

阅读类:包括教科书、专著、文集、科普读物等。教科书一般只介绍基础知识和公认的见解。科学专著是专门就某一课题或研究对象进行比较全面深入论述的学术性著作。文集是由各种文章(论文、报告等)汇编而成的一种出版物。

参考工具书:包括字典、辞典、指南、人名录、机构指南、手册、年鉴、百科全书等,这些参考书的内容可能是数据、表格、图解,也可能是文章条目,按一定的顺序排列,并有详细的索引。

正式出版的图书,均冠有 ISBN 号,即国际标准书号(International Standard Book Number),其定长为 10 位数字,分为 4 个部分,如 ISBN7-81034-660-1,其中“7”是地域号(国家、地区、语言区)指中国,若为 0 或 1 指英语国家,2 指法语区,3 指德语区,4 指日语等;“81034”是出版社号;“660”是书号;“1”是计算机校验号。ISBN 号前 3 部分是变长结构,但总长不变。我国出版的图书在 ISBN 号之后,还有一个《中国图书馆分类法》(简称中图法)的大类类号,便于图书排架和收藏。

2. 科技期刊 (journal, periodical) 一般是指采用统一名称(刊名)定期或不定期出版的连续性出版物(serials),有连续的卷、期号或年、月顺序号。期刊在内容上大都由单篇论文组成,各有专题,互不联系。广义的期刊包括杂志、报纸、年度报告、年鉴、丛书以及学会的会议录、学报和纪要等。据调查,科技人员从期刊中获得的科技情报,约占其整个情报来源的 70%,因此它是主要的情报信息源。

与图书相比,期刊的出版周期短,报道速度快,数量大,学科广泛,内容新颖,流通面广,能及时反映世界科技发展水平,因而在科技文献中占有非常重要的地位,也是科技人员为获取最新信息而经常使用的一种出版物,同样也是主要的检索对象。

同图书一样,正式出版的期刊也有国际标准连续性出版物编号 ISSN (International Standard Series Number),例如 ISSN1091-6490,实现对全世界期刊的管理。ISSN 号全长 8 位,前 7 位是刊名代号,末位是计算机校验号。中间的“-”只是为了便于阅读而设置的。

核心期刊 (core journals) 是科技期刊中一类特定的期刊,指刊载某学科文献密度大、载文率、被引用率及利用率较高,深受本学科专家和读者关注的期刊。

3. 资料 (material) 非书非刊的出版物,又称特种文献。它通常指那些在出版发行方面或获取途径方面比较特殊的文献,或难于搜求,或具有某些解决纠纷意义的文献。

(1) 专利文献:是一种用法律形式来保护的文献,包括专利说明书、权利要求书、摘要、附图等,核心是专利说明书。专利文献涉及的技术内容广泛,比较具体可靠,能较快地反映世界各国科学技术的发展水平,有统一的格式,文字简练。

(2) 科技会议文献：是科技工作者在各种学术会议上，交流科研新成果、新进展及发展趋势的讨论记录或论文等。科技会议文献内容新颖，学术性强，它不仅是提供信息的重要来源，也是迅速获得最新技术信息的一种重要途径。科学技术的新发现、新成果和新见解很多是在学术会议上首先公布的。

(3) 科技报告：是关于某项科学研究成果的正式报告，或是对研究和试验过程中各阶段进展情况的实际记录。其内容比较专深具体，很多研究课题及尖端学科的资料，往往首先反映在报告中，它不仅叙述了成功的经验，也记载了失败的过程，因此对科研人员起到直接借鉴的作用。科技报告能代表一个国家和专业的发展水平与动向，是不定期出版物，一个报告为一个单行本，有统一编号。

世界上每年的科技报告达数十万件之多，其中较著名的有美国四大报告：AD（美国武装部队技术情报服务处）报告；PB（美国出版局）报告；NASA（美国国家航空与宇航局）报告；NEC/ERDA/DOE（美国原子能委员会/能源研究与发展署/能源部）报告。

(4) 政府出版物：是指各国政府部门所发表并由政府专设机构统一出版的文献。该文献对了解某一国家的科技政策、经济政策及其演变情况，有一定的参考价值。政府出版物内容广泛，数量特别多，部分文献在列入政府出版物之前，已被所在单位出版过（如科技报告），所以有一定的重复。

(5) 学位论文：是高等院校研究生、本科生所写的作为评定学位的论文。用以介绍他们的研究成果和所得结论的调查报告。学位论文具有独创性，内容专一，论述详细、系统，尤其是博士论文有一定的参考价值，是经过答辩和评审的原始研究成果。大多数学位论文未进入出版发行渠道，不易被人们所利用。

(6) 标准文献：是对工农业产品和工程建设的质量、规格及其检验方法所做的技术规定，是从事生产、建设的一种共同技术依据，有一定的法律约束力。它是独立、完整的科技文献，有严格的审批程序，是各方专家集体制定的，内容可靠，其技术信息可直接使用。标准文献的新陈代谢非常频繁，随着经济条件与技术条件的改进，需经常不断地进行修改和补充。

(7) 产品资料：是指各国厂商对定型产品的性能、构造、原理、用途、使用方法、产品规格等所做的具体说明，又称产品目录、产品样本和产品说明书。该文献是生产科研单位研究分析各国技术发展和产品水平的重要资料，介绍的技术较成熟，是科技人员选型、设计和引进国外设备仪器有价值的参考资料。

(8) 技术档案：是指生产建设和科学技术部门在技术活动中所形成的，有一定具体工程对象的技术文件、图样、图表、图片、原始记录和原本以及代替原本的复制本。是生产过程及科技研究工作中用以积累经验，吸取教训，提高质量的重要文献，对科技人员有重要的使用价值。技术档案是某一工程项目的完整、忠实的记录，内容详尽、具体，是科技储备的一种最高、最完善的形式，它一般有密级限制，借阅手续严格。

三、科技文献的特点

随着科学技术的迅速发展，科技文献在形式、内容、出版、发行等各方面都表现出了许多新的特点。主要表现在以下几个方面。

（一）数量多、增长快

随着科学技术的不断向前推进，科技成果大量涌现，科技交流广泛开展，使科技文献的数量急剧增长。据联合国教科文组织统计，全世界每年出版的图书 80 万种以上，平均每 15

年翻一番，科技期刊在 10 万种以上；每年产生约 600 万篇文献，100 万件专利，其中生物医学文献占 30% ~ 40%。

（二）内容交叉渗透、分散重复

由于科学门类发展有既高度分化又高度综合的趋势，学科之间互相渗透，相互交叉，新的分支和边缘学科不断产生，致使文献分散、交叉出现。一种专业刊物所报道内容，往往包含了多个学科。从一个专业的文献而论，部分论文见于刊名为本专业的期刊，部分见于刊名与本专业相关的杂志上，还有部分见于刊名与本专业无关的杂志上。近年来，由于多种因素的影响，而出现一文多刊，一书多版，转载互译。一个专业或专题的文献，除期刊之外，常常又分散在专著、会议录、科技报告等之中。如此的交叉重复而又分散，无疑增加了搜集文献的难度。

（三）语种不断增加

20 世纪初，只要掌握英、德、法三种语言，就可阅读全世界 92% 以上的科技文献。而现在，发表科技文献的文种不断地趋于多样化。全世界的科技期刊涉及到的文种已达 70 ~ 80 种之多。如美国《医学索引》报道了 43 种语言文字的生物医学文献，美国《化学文摘》报道了 56 种语言文字的科技文献。文种的增加，造成了读者阅读文献的障碍，影响了情报信息的交流。

（四）失效快、寿命短

随着科学技术飞速发展，一项新技术，从理论到生产中应用、推广，其时间越来越短。反映在科技文献上的“新陈代谢”现象是文献的有效使用时间，这种时间称为“文献寿命”。由于文献类型不同，学科不同，其寿命也不同。科技图书的寿命相对来说要长一些，学位论文、专利文献等则要短一些。

（五）文献发表的时间滞后严重

科技文献增长的速度，要比科技文献发表的速度快得多，以致大量的文献不能及时发表，有的论文从编辑部收到稿件至正式发表之时，可长达一两年之久，一般文摘性刊物的报道时差也在几个月以上，致使一些文献被正式发表时就已失去了某些应有的价值，直接影响到文献的有效寿命和使用效果。为此，科技人员之间的直接交换手稿副本和预印本、通信、参观访问、参加会议、直接交谈等已成为非正式渠道获取未发表文献的有效途径。

（六）交流传播速度加快

现代交通、通讯、光学和印刷技术的发展，特别是现代通讯技术和电子计算机相结合应用技术的飞速发展，以及情报信息载体的电子化等在信息交流传播中的应用，为文献情报信息的快速传递与交流提供了非常便利的条件，文献情报信息的用户，可以通过信息交互网在瞬间即可获得所需要的文献情报信息。

（七）日益向多元化发展

科学技术的发展不断地使记载科技知识载体的形式多样化，且更具有科学性。目前，记录知识的载体已突破了传统的印刷型文献，缩微型、视听型、电子型文献也迅速发展起来。各种类型同时并存，各有长短，互相补充。

随着 Internet 网络技术的普及推广使用，从根本上改变了信息存取与传播的方式，电子型文献越来越受到人们的关注。数字化图书馆（digital library）、无纸社会（paperless society）等新概念由此而出现。计算机、网络将成为传播文化知识、获取信息资源的重要手段。

第三节 文献检索基本知识

面对日益庞大的、零乱分散的信息资源,人们如何从中获取自己所需的那部分信息,如何利用科学的方法来快速、准确、全面地收集和查找在科研活动中的重要文献信息,是文献检索所要解决的问题。

一、文献检索的概念、原理和意义

(一) 文献检索的概念

1. 文献检索的定义 文献检索是将文献按照一定方式集中组织和存储起来,并按照文献用户需求查找出有关文献或文献中包含的信息内容的过程。所以广义上的文献检索实质上是包括文献的存储和文献的检索两个过程。而我们通常所说的文献检索就是找出自己所需的文献,这只是对文献检索狭义上的理解。文献的存储和文献的检索两过程是相辅相成、密不可分的。所以文献检索的全称又叫“文献存储和检索”。

2. 文献检索的类型和范畴 一般认为,信息检索就是情报检索(information search),而作为检索对象的情报,它又有不同的存在形式。根据检索对象的不同,情报检索又分为文献检索、数据检索和事实检索。

(1) 文献检索:凡以文献为检索对象的,叫做文献检索。从性质上说,文献检索属一种相关性检索,即它只需检出与用户所需相关的文献供用户参考。例如:检索“有关生态关系社会的发展与人类健康方面”的文献,就是属于文献检索范畴的问题。

(2) 数据检索:以数据为检索对象的,叫做数据检索。它是一种确定性检索,即它能够直接提供用户所需的确切的数据。例如:查找“9·11 纽约世贸大厦被炸中死亡人数是多少?”就是一种确定的数据检索。

(3) 事实检索:以事实为检索对象的,叫做事实检索,它也是一种确定性检索。例如:查找“在 2001 年世界上有哪几个国家、哪些实验室在研究炭疽菌?”则是一种确定的事实检索。

但是文献检索、数据检索和事实检索三者并没有严格的界限划分,文献检索这个用语随着社会信息化进程的加快,其内涵也更为丰富和广泛。有观点认为广义的文献检索与情报检索的意义相同,根据课题的特定要求,查找出课题所需的特定文献、数据和事实的整个过程都是文献检索的范畴。

文献检索一般按照检出结果的形式不同可划分为以下几种类型:

(1) 书目检索:指检索出来的结果是关于文献的题录信息,即获得的是关于文献的题名、作者、摘要、出处等此类的文献线索信息,检索者可根据这些文献线索进一步获取文献原文。如:美国《医学索引》、荷兰《医学文摘》等都属于书目检索工具。

(2) 全文检索:是指检索出的结果直接为文献的全文内容,或者是按照用户需求,检索出有关的句子、章节或段落。全文检索是在计算机检索技术的支持上建立起来的。如:《中国期刊网全文专题数据库》、《ProQuest 保健、医学与药学全文数据库》等,都是全文检索。

(3) 引文分析检索:指检出结果为文献被引用的情况,主要用以分析和评价文献质量。如《科学引文索引》(Science Citation Index, SCI),可检索出某著者或文献被其他著者或文献引用的情况;又如《期刊引文分析数据库》(Journal Citation Reports, JCR)是一种关于期刊

评估的基本的、全面的资源工具,可查询期刊的影响因子 (impact factor, IF), 可以根据期刊引用数据反映期刊的重要程度。

3. 文献检索的方式 一般分为手工检索和计算机检索。手工检索方式是指利用印刷本检索工具进行人工查阅,这类检索一般不能实现全文检索服务。计算机检索是指利用计算机、激光或磁存储器以及现代化通信手段所实现的检索方式,如:光盘检索、联机检索和网络检索,这种检索方式为全文检索服务提供了有利条件。计算机检索以其数据更新快、检索速度快和智能化多途径检索等诸多优点,已成为现阶段文献检索的主要方式,但是计算机检索要完全替代手工检索还需要一段时间。计算机检索与手工检索的原理是一致的,在学习有关检索语言、索引技术和信息资源方面的知识时,手检比机检更容易理解。所以了解手工检索知识是很有必要的,是掌握计算机检索的必要基础。

(二) 文献检索的原理

文献检索包括文献的存储和检索这两个过程,它们的实现有赖于检索工具的存在。文献的存储过程就是检索工具的编制过程,这一过程通常由信息工作者去完成;而文献的检索过程则是检索工具的利用过程,可由文献需求者或情报人员去完成。所以任何检索工具都具有存储和检索两方面的职能。存储是检索的基础,而检索是实现存储的目的和手段,两者相辅相成、互为依存。文献检索的核心就是要使文献的存储与检索两过程所采用的特征标识达到一致。

存储过程,就是形成文献特征标识,为检索提供有规可循的检索途径的过程。该过程首先是要为每一篇纳入检索工具(或系统)的文献标记上检索标识,即对文献进行全面分析,从自然语言转换成规范化的检索语言的过程,这也就是文献的标引过程。检索标识有两种:一种是文献的外表特征,也称为自然标识,包括标题、著者、来源、卷期、页码、年月、类型、号码、文种等项目;另一种是文献的内容特征,也称为人为标识,主要是文献论述的主题,亦即文献的中心内容,诸如主题词、分类号、类目名称、文摘等。标引内容特征时,首先需要对原始文献的内容进行主题分析、深入理解,找出若干能代表文献中心内容的概念,然后用情报检索语言的语词(或标识)来把这些概念标识出来;如对文献给予分类号的过程称为分类标引,给予主题词标识的过程称为主题标引。其次是要将有关文献特征标识著录下来,形成一条条有规可循的文献线索。最后将这些文献线索系统地按一定规则排列到检索工具(或系统)中去。

检索过程,就是利用特征标识,找出相关文献的过程。该过程首先是要对检索课题进行主题分析,找出若干能代表课题需要的概念,即确定主题检索概念;然后用存储过程中所使用的情报检索语言,将这些主题检索概念转换成检索特征标识;最后利用这些特征标识到检索工具(或系统)中去查找到相关文献线索。

由此可知,检索过程是与存储过程相对应的逆过程,而连接两个过程的纽带就是情报检索语言。文献检索的基本原理就是检索者将主题检索概念与存储在检索工具(或系统)中的检索特征标识进行比较,最后达到一致,以此来有效地查得并获取文献。

(三) 文献检索的目的和意义

1. 继承和借鉴前人成果,避免重复研究 科学技术的发展具有连续性和继承性的特点。科学的每一进展都是在前人知识的基础上取得的,科学知识的每一次进步都是对某一方面的补充、修改和精练。对于一个科研工作者来说,文献信息检索是其科学研究不可缺少的一部分,在其着手开展一个新的研究课题之前,必须全面了解这个课题的背景资料,如前人在这方面已做了些什么工作,是怎么做的,还存在什么问题等等。只有准确地搜集了相关资料,摸清

了该科学技术发展的水平、动向,吸取已有的科技成果,才能避免科研工作中重复浪费,少走弯路。例如美国 20 世纪 50 年代,为了做“继电器接点电路合成研究”,曾经联合几家实验室研究了 5 年,耗资 50 万美元终于完成,但当发表成果时才发现该项目早已有其他人完成。又如美国某轧钢厂一位化学家,花了 1 万美元完成了一系列实验,并解决了问题,而正当他为自己的成果得意时,却得知有一份德国人的报告与其实验完全相同,全部资料只需花 5 美元就行了。这些例子都说明,如果在课题研究之前,能够先检索有关文献,了解其当前发展水平,就一定能避免重复劳动,并扬长避短,开拓思路,在一个新的起点上作出努力。

2. 节省科研工作时间,提高科研效率 文献的庞大数量和迅速增长,以及文献的分散、交叉重复加重了研究人员搜索信息的负担。许多科学家和信息学家一再表明,不会获取信息将对科学研究造成损失和失败;不会获取有用的信息,将让科研人员在五光十色的信息海洋中感到困惑和无奈。但是如果懂得文献检索的方法和技巧,尤其是以计算机检索为主体的现代化信息检索技巧,就可以比较容易检索到针对性较强的文献,达到事半功倍的效果。

3. 协助决策者做出正确决策 所谓决策,就是指为了实现某一特定目标,从客观实际出发,在占有一定信息和经验基础上借助于一定的方法,从各种可供选择的方案中选出最佳方案的活动。正确的决策虽然与多种因素有关,如决策体制、决策方法、决策者的能力和素质等,但是决定性的因素还是决策者对客观实际的了解,对未来行动和后果的正确判断等,而所有这些主要依赖于全面、即时、准确的信息保障。“知己知彼,百战不殆”,所以说,情报是决策的重要依据。

4. 终身学习、知识更新的必备工具 时代在发展,信息在瞬变,科学技术日新月异,知识更新周期缩短,尤其以计算机技术和卫星通讯技术为主的现代信息技术更是变幻莫测。有人说,70 年代的大学本科毕业生就业后其知识可以用 8 年,80 年代可以用 5 年,而 90 年代只能用两年。21 世纪,知识的新陈代谢更是加快,学习、吸收和应用新知识、新信息、新技术的能力已构成竞争力的关键因素。与此同时,为了迎接知识不断更新所带来的挑战,人们需要从在校学习转化为终身学习,而终身学习、知识更新都依赖于信息的获取和利用。

5. 开发情报信息资源,提高经济效益 近十年来,世界在发生翻天覆地的变化,在目不暇接的经济万花筒背后,一种新的经济形式——知识经济正向我们席卷而来,它正将人类社会带向 21 世纪知识经济时代。知识经济是社会发展的必然。早在 20 世纪 70 年代,美国未来学者托夫勒在其著作《第三次浪潮》中就指出了人类将面临信息社会的到来,此后,经济学家们开始关注这一提法。回顾近半个世纪以来的世界经济发展状况,50 年代西德重新崛起,60 年代日本经济奇迹般增长,70 年代出现亚洲四小龙,进入 80 年代停滞徘徊多年的美国经济也开始复苏并得到发展,这些发展都是以知识为基础的。1990 年联合国研究专家正式提出了“知识经济”这一概念,知识成为发展经济的最重要的直接资源,而信息成为一个国家生存和发展的关键所在,成为国际竞争中的重要因素,控制了信息,尤其是高新尖端的科技信息,也就获得了经济发展的主动权和国际竞争的优势。在这样一个社会经济概念面前,世界各国、各民族、各行业都在积极应对。谁能有效地捕捉信息,分析研究信息并加以利用,谁便能在强手如林、变幻莫测的市场上抢占有利的位置。

二、检索语言

文献检索的原理是要使文献的存储与检索两过程所采用的特征标识达到一致,而使其一致的中间纽带就是检索语言。检索语言是检索者与检索工具之间进行交流的标准,也是情报

存储人员与检索工具之间交流的标准。检索者、检索工具、情报存储人员三者通过检索语言来互相沟通。

(一) 检索语言的概念

检索语言 (retrieval language) 是用于描述检索系统中文献的内容特征及外表特征和表达用户检索提问的一种专门语言。检索的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。检索效率的高低,在很大程度上取决于所采用的检索语言的质量以及对它的使用是否正确。

(二) 检索语言的类型及特点

检索语言由检索标目和语法组成。检索标目是指登录在检索系统或工具中的具有检索意义的文献特征标识,语法则是指利用这些标目查找文献的一整套规则。检索语言有很多种,按照不同的标准可有不同的划分,归纳起来主要有以下两种分法。

1. 按照检索语言中所使用语词的受控情况可分为规范语言 (controlled language) 和非规范语言 (uncontrolled language)。

规范语言,又叫受控语言、人工语言 (artificial language),是一种采用经过人工控制了规范性的词语或符号作为检索标目,来专指或网罗相应的概念,这些规范化的标目能较好地同义词、近义词、相关词、多义词及缩略词等概念进行规范。比如“肿瘤”这个概念的英文词汇中有多个同义词: tumor、tumour、carcinoma、neoplasms 等,人工语言则规定只能用其中一个词来表示所有“肿瘤”的概念,假如人为规定的词是“neoplasms”,那么无论原始文献中作者使用的是哪一个词,检索者只需用“neoplasms”就能将所有肿瘤的文献查找出来。因此,使用规范化的词能相对提高检索的效率,但对检索者和情报存储人员在选词上要求比较严格。在后面所述的单元词语言、标题词语言、叙词语言等都是规范性语言。

非规范语言,又叫非受控语言、自然语言 (natural language),是采用未经人工控制的词语或符号来作为检索标目,通常所说的自由词、关键词就属此类。一般当某些特定概念无法用规范词准确表达,或新出现的词语(如纳米)还未来得及被规范化时,都需要使用非规范词。这类词语有较大弹性和灵活性,检索者可以自拟词语进行检索。在机检中,非规范词的应用比较广泛。但这类语言对一词多义、多词一义的词语,检索就相对困难些。在后面所述的关键词语言就是一种非规范语言。

2. 按照检索语言所描述的文献信息特征的内容来分,其类型如图 1-1 所示。

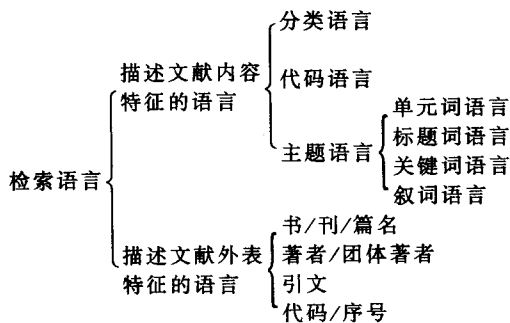


图 1-1 检索语言的类型

不同的检索语言构成不同的标识和索引系统,提供用户不同的检索点和检索途径。

描述文献外表特征的检索语言，是依据文献外表特征作为文献存储的标识和文献检索提问的出发点而设计的索引语言，如以文献上记载的书名、刊名、篇名等作为标目的书名目录（索引）、刊名目录（索引），以及“引用期刊一览表”、篇名索引等；以文献上署名的著者、译者、编者等姓名或机关团体名称作为标识的著者目录（索引）、专利权人索引等；还有以文献末尾所附的参考文献或引用文献的特征作为标识的，如引文索引；以及以文献特有的序号或代码作为标识的索引系统，如专利号索引、科技报告号索引等，这些都是采用描述文献外表特征的检索语言建立起来的。

描述文献内容特征的检索语言，按其构成原理分为分类语言、代码语言和主题语言。

（1）分类语言：是一种直接体现知识分类的等级制概念标识系统，是按照学科范畴划分而构成的一种语言体系。它能集中体现学科的系统性，反映事物的从属、派生关系，并从上至下、从总体到局部层层划分、展开，是一种等级分明的语言。这种语言一般由类目（语言文字）或其相对应的类号（字母、数字或它们的组合）来表达各种概念，构成一个完整的分类类目表，如《中国图书馆分类法》、《国际专利分类表》等都是典型的分类语言。

（2）代码语言：是对文献所论述事物的某一方面的特征，用代码加以描述和标引，如：化学物质的分子式、化学物质登记号、基因符号等。

（3）主题语言：根据其表达概念的不同形式又分为关键词语言、标题词语言、单元词语言和叙词语言。

1) 关键词语言 (keyword)：是由直接从文献的篇名、文摘或全文中抽取出来的词汇构成的。在检索工具中常以关键词索引出现。这些词汇都对表达文献主题具有一定的实质意义，几乎所有有意义的信息单元都可以作为关键词，而那些对文献主题表达无意义的词汇，如冠词、介词、连词、副词、感叹词，以及某些形容词、某些动词（系动词、情态动词、助动词）等，均不能作为关键词，并被归入非关键词表或禁用词表。关键词语言基本上是未经规范化处理的语言，属于自然语言的一种，但实际运用中仍对少量词作了规范化处理，如复合词的拆散和组合，词和符号的改写等。关键词语言具有易于掌握，方便检索，灵活性高，在计算机检索中应用广泛等诸多优点，但它未经规范，用词不统一，因而漏检率高，检全率较低。

2) 标题词语言 (heading)：是最早使用的一种主题词语言，它使用的词汇也是从文献内容中抽出，但与关键词语言不同的是，它们都是能代表文献内容特征的规范化词语，属于规范性语言的一种。

标题词分为主标题词和副标题词。标题词表中出现的这两级标题词以固定的方式组合在词表中，检索按既定组配进行。因而使用标题词语言检索之前，词与词之间的组配已经固定，这种形式又称为先组式语言。标题词语言属于规范性语言，因而具有较好的通用性、直接性和专指性，但其又是先组式语言，因此灵活性较差。

3) 单元词语言 (uniterm)：是一种最基本的、不能再分的单元词语，亦称元词。它也是从文献内容中抽出，并经过规范化处理的。但它仅能作为元词（或单元词），代表一个独立的概念，比如“肝功能”不是单元词，只有“肝”和“功能”才是单元词。在英语中元词经常是一个单词。元词具有灵活的组配功能，在检索时可以将某些元词组配起来代表某一特定概念，因此这种语言属于后组式语言，如《世界专利索引》工具中的《WPI-规范化主题词表》就属于单元词语言。

4) 叙词语言 (thesaurus)：叙词是在标题词、单元词和关键词基础上发展起来的，它和