

21世纪高等医学院校教材

现代医学

信息检索

主编 李健康 夏旭

海天出版社

21世纪高等医学院校教材

现代医学信息检索

主 编:李健康 夏 旭

副主编:周 纯 朱培毅 李子丰 刘爱山

参编单位及编委会成员:(按汉语拼音字顺排序)

广东医学院 陈文凯 管 进 李子丰 万 冰

广州医学院 朱培毅

南方医科大学 顾 萍 李健康 李 梅 夏 旭

许四洋 续向军 张春辉 张政宝

汕头大学医学院 刘爱山 余恩琳

中山大学 周 纯 黄晴珊 罗淑莲

图书在版编目(CIP)数据

现代医学信息检索 / 李健康, 夏旭主编, —深圳: 海天出版社, 2005. 5

21 世纪高等医学院校教材

ISBN 7-80654-376-7

I. 现... II. ①李... ②夏... III. 医药学—情报检索—医学院校—教材 IV. G252.7

中国版本图书馆 CIP 数据核字 (2005) 第 025005 号

海天出版社出版发行

(深圳市彩田南路海天综合大厦 518033)

<http://www.htpb.com.cn>

责任编辑: 张悟颖 封面设计: 亮点设计工作室

广州市番禺新华印刷有限公司印刷

海天出版社经销

2005 年 5 月第 1 版

2005 年 5 月第 1 次印刷

开本: 787mm×1092mm

1/16

印张: 20

字数: 500 千

定价: 38.00 元

海天版图书版权所有 侵权必究

内容简介

本书根据 21 世纪现代医学信息检索的突出特点,融合了传统文献检索和现代信息检索的基础理论;以信息检索、计算机检索、互联网及搜索引擎知识为基础,系统阐述了常用中外文生物医学数据库、全文数据库和引文检索的特点与使用方法;以数据库检索为重点,以搜索引擎为辅助手段,全面介绍了互联网医学信息检索的方法和特点;对于 PubMed 等网上虚拟数据库的检索、会议预报信息和专利信息的网上获取技术、使用电子邮件和电子论坛发布消息、网上会议信息发布、以及核酸序列的提交等,都做了比较全面的介绍。全书内容简明新颖,实用性强。既适合于医学本科生、研究生获取网上生物医学信息之用,也适合于医学临床、教学和科研人员查询网上信息时参考,还可作为图书信息人员案头必备参考书。

前言

互联网的发展和网络资源的快速增长,使得信息资源的有序组织与检索利用正成为网络资源开发利用的重要课题。《文献检索与利用》课正是基于学生的信息素质和信息能力培养而开设的一门方法课。经过 20 余年的教学实践,已经演化为提升学生的素质特别是信息素质的重要课程,成为掌握读书治学门径、提高信息资源开发利用能力的重要工具。在医科大学,该课程已经普及到博士、硕士、本科、大专和成人教育等所有层次。围绕信息素质教育的开展,国内出版了不少以培养信息意识和信息能力为目的的文献信息检索教材。这些教材增加了计算机检索和网络信息检索内容,实现了从传统的文献检索向现代信息检索的快速转变。

《现代医学信息检索》一书的出版顺应了广大医学生信息意识与信息能力培养的要求。它立足于当前高校文献检索教学、参考咨询、情报调研和科技查新的具体实践,深刻而全面地反映了现代信息检索领域的新理论、新方法、新成果,力求内容科学新颖、简洁实用,帮助学生系统而全面地掌握现代医学信息检索的核心技术与主要方法,打开通向网络信息资源检索之门。

该书为 21 世纪高等医学院校教材,以广东省 5 所高等医学院校的文献检索课教学实践为基础,集中了这些高校中从事文献检索课教学与研究的中坚力量,囊括了各个医学院校的教学内容,同时吸收了国内文献检索课教学的有益经验和相关教材内容的精华,从而保证了教材的编写既符合信息素质教育与培养的要求,又切合各个学校的开课需要。该书的编者都是从事多年文献检索课教学的老师,在这方面具有丰富的经验。他们既是现代医学信息检索的实践者,又是信息检索教学的传道授业解惑者,书中所述各章节内容无不是他们的教学经验的总结和教学成果的体现。

全书共分十三章。第一章绪论,介绍了信息、知识、情报、文献的定义及其相互关系,文献信息检索方法、技术、途径与步骤,创造性地整合了网络数字型文献及网络信息检索工具。第二、三、四章介绍了国内外常用的生物医学数据库、期刊全文数据库和美国生物技术信息中心(NCBI) Entrez 数据库的特点与检索方法。第五—八章分别介绍了引文检索、医学会议信息检索、医学专利文献检索和学位论文数据库检索。第九章重点阐述了文献全文获取的方法途径步骤。第十章概述了循证医学、循证医学证据与循证医学证据检索,论述了循证医学证据资源检索特点与检索方法。第十一章概述了搜索引擎的定义、工作原理与基本结构,重点介绍了中外通用与医学搜索引擎。第十二章介绍了网上医学信息发布的几种主要方法:利用电子邮件、向 BBS 论坛及新闻组发送信息、预报会议信息的网上发布、网上核酸序列提交。第十三章从概述、医学信息的搜集、鉴别与整理、医学信息分析方法、医学信息调研成果的撰写等方面介绍了医学信息调研的主要内容。

与同类教材相比,本书在内容的选择上注重了学术性、实用性、创新性与可获得性,同时也考虑了不同学校 and 不同单位的实际情况,力求内容典型、精练、新颖,具有代表性和可操作性。

性。如第一章绪论,创造性地将数字化文献资源和网络信息检索工具融进了传统文献检索教材所涉及的文献类型和检索工具分类;第十一章搜索引擎介绍了学术性、专业性很强的科学搜索引擎 SCIRUS;第十三章网上医学信息发布提供了比较全面而权威的免费电子邮箱地址和网上论坛列表,这些都是同类书籍中鲜有介绍的新内容。

本书在编撰过程中参阅了国内外大量文献检索和文献检索教学论文、专著和网络资源,在此,向这些作品的作者们一并致谢。

由于网络资源的瞬息万变、网络检索工具和数据库的不断更新、功能不断加强,作者们虽执著于用例新颖、言简意赅,然而编撰内容仍无法跟上网络的发展变化进程;又,编者们的知识背景、社会阅历、写作水平、行文风格、驾驭文字的功夫俱有差异,故问题和错误在所难免,敬祈读者斧正。

编 者

2005年4月12日

目 录

第一章 绪 论	1
第一节 信息、知识、情报、文献及其相互关系	1
第二节 现代医学文献信息的类型与发展特点	4
第三节 医学文献信息检索工具	9
第四节 文献检索与信息检索	11
第五节 检索语言及其应用	15
第六节 检索策略与检索技术	26
第二章 常用中文生物医学数据库检索	37
第一节 中国生物医学文献数据库	37
第二节 中文生物医学期刊数据库	50
第三节 中国期刊全文数据库	56
第四节 中文科技期刊数据库	65
第五节 图书全文检索	68
第三章 常用外文生物医学数据库检索	77
第一节 Medline 光盘数据库	77
第二节 EBSCOhost 期刊全文数据库	83
第三节 SpringerLink 期刊全文数据库	90
第四节 ProQuest Mediacal Library 期刊全文数据库	92
第五节 Science Direct Onsite 期刊全文数据库	96
第六节 BIOSIS Previews 数据库	99
第七节 Web of Science 数据库	107
第八节 CA on CD 数据库	112
第九节 TOXNET 数据库	116
第十节 CancerLit 数据库	122
第十一节 Ei Compendex 光盘数据库	124

第四章	NCBI Entrez 数据库检索	132
第一节	生物信息学与 Entrez 系统	132
第二节	PubMed	138
第三节	核酸序列数据库	146
第四节	蛋白质序列数据库	151
第五节	蛋白质结构数据库	152
第六节	基因组数据库	154
第七节	OMIM 孟德尔人类遗传数据库	155
第五章	引文检索	159
第一节	中国期刊全文数据库	159
第二节	中文科技期刊引文数据库	163
第三节	中国生物医学期刊引文数据库	168
第四节	Web of Science	172
第六章	医学会议信息检索	180
第一节	国内医学会议信息检索	180
第二节	国外医学会议信息检索	184
第三节	医学会议信息相关网站	187
第七章	医学专利文献检索	191
第一节	专利与专利文献	191
第二节	中国专利数据库	194
第三节	国外专利数据库	199
第八章	学位论文数据库检索	208
第一节	国内学位论文数据库	208
第二节	国外学位论文数据库	216
第九章	文献全文获取	221
第一节	途径与方法	221
第二节	免费网络医学期刊	223
第三节	免费网络医学图书	226
第四节	网络文献传递服务	229
第十章	循证医学证据检索	234
第一节	循证医学概述	234
第二节	循证医学证据概述	237
第三节	循证医学证据检索概述	241

第四节 循证医学证据资源检索特点	244
第十一章 搜索引擎	253
第一节 概述	253
第二节 通用搜索引擎	254
第三节 生物医学搜索引擎	264
第十二章 网上医学信息发布	275
第一节 利用电子邮件发送信息	276
第二节 向 BBS 论坛及新闻组发送信息	282
第三节 预报会议信息的网上发布	291
第四节 网上核酸序列提交	293
第十三章 医学信息调研	295
第一节 医学信息调研概论	295
第二节 医学信息的搜集、鉴别与整理	297
第三节 医学信息分析方法	302
第四节 医学信息调研成果的撰写	304
参考文献	308

第一章 绪论

第一节 信息、知识、情报、文献及其相互关系

一、信 息 (information)

“信息”一词的释义众说不一。从哲学角度可概述为,信息是物质的一种带有普遍性的关系属性,是物质存在方式及其运动规律、特点的外在表现。从信息传递角度来认识,信息则是关于自然界、生物界和人类社会中一切事物运动状态及关于事物运动状态的报导。总之,信息是自然界、人类社会及人类思维活动中存在和发生的一切宏观和微观现象,大至天体、小至细胞、原子、电子、基本粒子等现象,故一切消息、知识、数据、文字、程序和情报等都是信息。信息是事物的运动状态与方式的反映,是生物体或有一定功能的机器通过感觉器官或设备与外界交换的内容。不同的事物有不同的运动状态与方式,因而会产生不同的信息。

为了研究和使用的需要,可按不同的标准对各种信息进行归类:

按内容分有:①状态信息。指关于事物的状态、属性等存在方式。如花红草绿、天冷天热等,它们往往是从观察实验中获得的。②规律信息。反映事物的运动规律和某种定常性,如万有引力定律,太阳的东升西落等。由于科学理论一般都包含了规律性的描述,有时也可称之为理论信息、知识信息,它们是从对状态信息的动态分析中获得的。③指令信息。指行动的方向和方式,如向前走、停止等命令或指令。它是结合前两种信息并考虑到一定的目的性和约束条件以后得出来的。

按存在的领域来分有:①物质信息。它与一切物质的存在形态有关。②生物信息。指仅存在于生物界和人类社会之中的信息。

按形态品位来分有:①基本信息。与物质对时间和空间的分布与差异有关的信息。②交换信息。与物质的相互作用和关联有关的信息。③遗传信息。存在于生物遗传过程中。④演算信息。存在于算术和逻辑过程中信息。⑤语言信息。⑥情感信息。

按信息的属性来分有:①文字信息。文字是人们为了实现信息交流、通信联系所创造的一种约定的形象符号。广义的文字还包括各种编码,如 ASCII 码、汉字双字节代码、国际电报与单元代码以及计算机中的二进制数字编码等。②图像信息。图像是一种视觉信息,它比文字信息直接,易于理解。人工创造的图像,如一张纸、一幅画、一部电影、大自然的客观景象等都是抽象或间接的图像信息。随着多媒体技术的发展,各类图像信息库将会极大地丰富人类的生活。③数值数据信息。数值数据是“信息的数字形式”或“数字化的信息形式”。狭义的数据是指有一定数字特性的信息,如统计数据、气象数据、测量数据以及计算机中区别于程序的计算数据。广义的数据是指在计算机网络中存贮、处理、传输的二进制数字编码。文字信

息、图像信息、语音信息以及从自然界直接采集的各种自然信息均可转换为二进制数码,网络中的数据通信、数据处理和数据库等就是广义的数值数据信息。④语音信息。人讲话实际上是大脑的某种编码形式的信息转换成语音信息的输出,是一种最普遍的信息表现形式。音乐也是一种信息形式,是一种特殊的声音信息,它是通过演奏方式表达出来的丰富多彩的信息内容。

所谓医学信息,就是用语言、文字、符号、图像、影视数据等反映人类与疾病做斗争的过程中疾病运动状态及其变化方式的信息。

信息是一种宝贵资源,它同物质、资源并列,共同推动人类社会向前发展。信息的显著特征是以物质为其存在载体的资源,可进行客观处理和传递,因而具有共享性和开发利用的无限性。但对信息的开发利用很有针对性,与人们情报意识的强弱有着密切关系。对于情报意识强者,信息是有用的,并可以增值;对情报意识弱者,则视而不见,听而不闻,完全失去作用。

二、知识 (knowledge)

知识历来是哲学中认识论研究的对象,故常见的有关知识的定义是从哲学的角度提出的。我国《教育大词典》中流行的知识定义是:“对事物属性与联系的认识。表现为对事物的知觉、表象、概念、法则等心理形式。”或者《中国大百科全书·教育卷》更具体:“所谓知识,就它反映的内容而言,是客观事物的属性和联系的反映,是客观世界在人脑中的主观映象。就它反映活动的形式而言,有时表现为主体对事物的感性知觉或表象,属于感性知识,有时表现为关于事物的概念或规律,属于理性知识。”这两个定义都是根据哲学认识论中的反映论给出的,强调知识是客观世界的主观反映。

当代最著名的认知心理学家皮亚杰认为:“知识是主体与环境或思维与客体相互交换而导致的知觉建构,知识不是客体的副本,也不是由主体决定的先验意识。”根据皮亚杰的思想和当代信息加工心理学的观点,可将知识定义为主体通过与其环境相互作用而获得的信息及其组织。贮存于个体内,即为个体的知识,贮存于个体外,即为人类的知识。

对知识概念表述确切而又容易被接受的是:知识是人们在改造世界的实践中所获得的认识和经验的总和,是人的大脑通过思维重新组合的系统化的信息集合。知识来源于信息,是对信息加工后获得的产品。

知识是信息的一部分,是人类的第二资源,具有指导与推动人们开发和利用第一资源(物质资源)的作用。

三、情报 (infotelligence)

无论中文还是英文,一直对“情报”存有争议。由于英文中只有 information 一词,译成中文既是“信息”,又是“情报”。长期以来由于没有更好的词来表述这两个概念,所以在已经出版的辞书中人们见到的都是用 information 来表述“信息”和“情报”两重意思。2001年12月1日-5日,中国国防科技信息学会情报学术专业委员会和中国科技情报学会理论方法与教育专业委员会在湖北省老河口市联合召开的“新世纪初情报学科建设、发展与应用研讨

会”会上再次重申,科技情报范畴中的汉语“情报”术语在英文中,既不是 information,也不是 intelligence,而是二者的整合 infotelligence。

情报的最初含义是指那些时间性很强的消息传递,即把情报视为一种活动,如军事、谍报工作。随着科学技术的发展,情报概念的内涵也随之发生变化,对其功能也有新的要求。如我国著名科学家钱学森指出:“情报就是为了解决一个特定问题而需要的知识。”“要注意它的及时性和针对性这个要求。”其内涵已扩展到人们对特定知识的需求。可见情报的本质是知识,具有传递性和实用性的特点。

什么是情报?情报就是为了解决一个特定的问题所需要的激活了、活化了的特殊知识或信息。一是知识,二是及时性和针对性。资料不是情报,信息和知识并不都是情报,智能也不是情报。情报的科学技术包括三个方面:一个是把有关的知识和信息收集起来;二是建立信息库,建立检索系统,以便于使用;三是把这些知识和信息活化、激活了以后,变成情报,也就是情报分析。情报分析工作要靠一个模型,把搜集到的点滴的资料输入到这个模型里去,这个全貌就出来了。如果没有模型,可以根据系统输入输出的数据,采用系统工程中系统分析方法对系统定型,再采用系统辨识的方法对系统定量,从而猜测出系统的模型。

四、文献(literature, document)

“文献”一词在不同的历史时期有着不同的内涵。有关“文献”概念的描述最早见于《论语·八佾》,朱熹解释为:“文,典籍也;献,贤也。”元代历史学家、文献学家马端临,首次以“文献”一词命名他自己的专著《文献通考》。他把“文献”概念的内涵表述为可作叙事、论事佐证的古今典籍及圣贤的言论,即文献兼指书和人。后来,文献概念的内涵偏指著述或专指“具有历史价值的图书文物资料”。1983年颁布的《中华人民共和国国家标准·文献著录总则》(GB3792.1—83)对文献概念定义为:“记录有知识的一切载体。”强调了文献是知识。凡是能耳闻、目睹到的并记录在一定的物质上的知识载体,均称之为文献。《文献情报术语国际标准》(ISO/DIS5127)将文献定义为:“在存贮、检索、利用或传递信息的过程中,可作为一个单元处理的,在载体内、载体上或依附载体内而存贮有信息或数据的载体。”强调了文献是可以进行加工处理的“单元知识”;存贮(记录)“单元知识”的方式、方法是多种多样的,如存贮声音、图像、文字、符号等知识的磁性载体(磁带、磁盘、光盘)文献等。

文献是“记录知识的一切载体”,这是“文献”概念的精辟定义。具体地说,文献是指文字、图像、符号、声频、视频等作为记录手段,将信息记录或描述在一定的物质载体上,有其特定的表现形式,并能起到存贮和传播信息情报和知识作用的一切载体。这种载体除了常见的纸质图书和期刊外,还包括胶片、磁盘、磁带、光盘以及甲骨、金石、竹帛等。由此定义可以看出,文献具有四个基本属性和三个基本功能,即:文献的知识性(内容)、记录性(手段)、感知性(媒介)和物质性(载体),存贮知识、传递和交流信息的功能。

文献是人类认识自然和社会过程的知识记录,是人类物质文明不断发展的产物,印下的一串串足迹可以再现历史、鉴往昭来;它凝结着数千年以来人类点滴积聚的智慧和成果,所蕴含的知识、信息、情报,是人们读书治学、开展科学研究活动和从事其它工作的智慧源泉,可以使知识增值;是启迪后人开创未来的丰富宝藏;它可以满足人们的文化需求,使人们从中获取有益的启迪与精神享受,是精神财富的重要组成部分。

五、信息、知识、情报与文献的关系

通过上述对信息、知识、情报与文献的分析,可以认为信息是宇宙间的一切运动状态及其报导。宇宙间时时刻刻都在产生着信息,人们正是通过对这些不同信息的获取来认识不同事物,并由此产生新的知识。知识是经人脑思维加工而成为有序化的人类信息。文献则是被物化了的知识记录,是被人们所认知并可进行长期管理的信息。情报是人们为解决特定问题而被活化了的更为高级、更为实用的知识。情报蕴含在文献之中,但不是所有文献都是情报,而所有情报几乎都是知识。文献又是贮存传递知识、情报和信息的介质,它们之间的逻辑关系是一种包含关系。

长期以来,人们习惯于从文献中获取信息、知识和情报,文献已成为人们获取信息、知识和情报的主要方式。因此,文献是一种重要的信息源。但是,文献并不是惟一的信息源。除了文献信息外,还有非文献信息源。非文献信息源又可分为实物信息源和口头信息源两种。前者包括实物、样品、展览等,后者包括交谈、会议、广播等。

第二节 现代医学文献信息的类型与发展特点

一、文献信息的级别

根据对文献内容的加工程度,可将文献分为多个等级。

1. 一次(级)文献

即原始文献。它是以生产或科研成果为依据而创造的原始文献。专著、期刊论文、研究报告、专利文献、学位论文、会议文献以及译文,数字文献资源中的事实数据库、电子期刊、电子图书、发布一次文献的学术网站等,均属于一次文献。

2. 二次(级)文献

对一次文献加工整序后的文献,包括目录、索引、文摘、参考数据库、网络学术资源学科导航、搜索引擎/分类指南等。二次文献是文献检索工具的主体。

3. 三次(级)文献

指为了一定的目的,通过二次文献利用一次文献进行再加工后编辑而成的文献。综述、年鉴、进展、指南、百科全书、科技动态、决策参考、情报调研报告等属三次文献。还有专门用于检索搜索引擎的搜索工具,比较典型的是 WebCrawler,被称之为“搜索引擎之搜索引擎”(search engine of search engine),即“元搜索引擎”,当用户进行检索时,反映出来的结果是各搜索引擎的检索结果。

4. 零次(级)文献

这是二十世纪七十年代提出的一个概念,经过长时间讨论得不到普遍认同,故不做具体介绍。

二、文献的类型

（一）印刷型文献

印刷型文献就是传统的纸质型文献,是指以印刷技术或手写为主要手段、以纸张为信息载体的文献。其优点是可以直接阅读、携带方便,是目前人类信息交流活动中最常用的工具。与现代信息载体相比,印刷型文献存储信息密度低,占用收藏空间大,不易长期保存,难以实现自动化输入和自动检索。

1. 图书(book)

一般包括专著、教科书、论文集和参考工具书。图书通常提供比较系统、成熟的知识。

2. 期刊(journal,periodical)

也称杂志,是指具有相对固定的刊名、编辑机构及版式装帧的连续出版物。期刊的内容通常是能够反映学科领域最新的理论、方法、技术的论文、综述、病例报告等。

3. 特种文献

又称非书非刊资料,即图书和期刊以外的出版物。其特点是数量大、种类多、内容广、参考价值大。特种文献包括:1) 政府出版物。多数是国家各级政府部门及其所属机构出版的行政文件,如各种法令法规、外交文件、统计数据等,科技资料数量较少。2) 会议文献。指国内外学术团体举行的专业会议上发表的论文或学术报告。其特点是信息传播速度快、反映研究成果新。常以论文集、会议论文摘要、期刊特辑或增刊的形式出现。3) 专利文献。指受法律保护的技术发明。专利文献是指专利发明人向知识产权局递交的、说明自己创造发明的技术文件,也是实现发明所有权的法律性文件,包括专利说明书、专利公报等。4) 科技报告。指各学术团体、科研机构、高等院校的研究报告及其研究过程的记录。其理论性较强,能反映某一专业领域的研究进展和动态,保密性强,通常难以获得。5) 技术标准和规范。指有关产品或工程质量、规格、生产过程、检验方法的技术文件,具有一定的法律约束力。6) 学位论文。指毕业生为取得博士、硕士、学士学位而撰写的学术性研究论文。7) 其它特种文献。包括手稿、内部刊物、报纸病历档案、技术资料、产品样本等。

（二）视听型文献

视听型文献也称声像型文献,是一种记录声音、图像信息的载体,如照片、录音带、录像带、幻灯片、影视片等。视听型文献的载体形式可以让人们通过视觉、听觉感受到直观、形象、生动、逼真、丰富多彩的信息。

（三）缩微型文献

缩微型文献是指以感光材料记录信息的一种载体,如缩微胶卷、缩微胶片、计算机存取载体的输出胶片等。缩微型文献载体体积小、存储密度高、成本低、便于保存,但使用时必须借助缩微阅读设备才能进行阅读或复印。

（四）数字型文献

1. 数字型文献的结构。

见图 1-2-1。

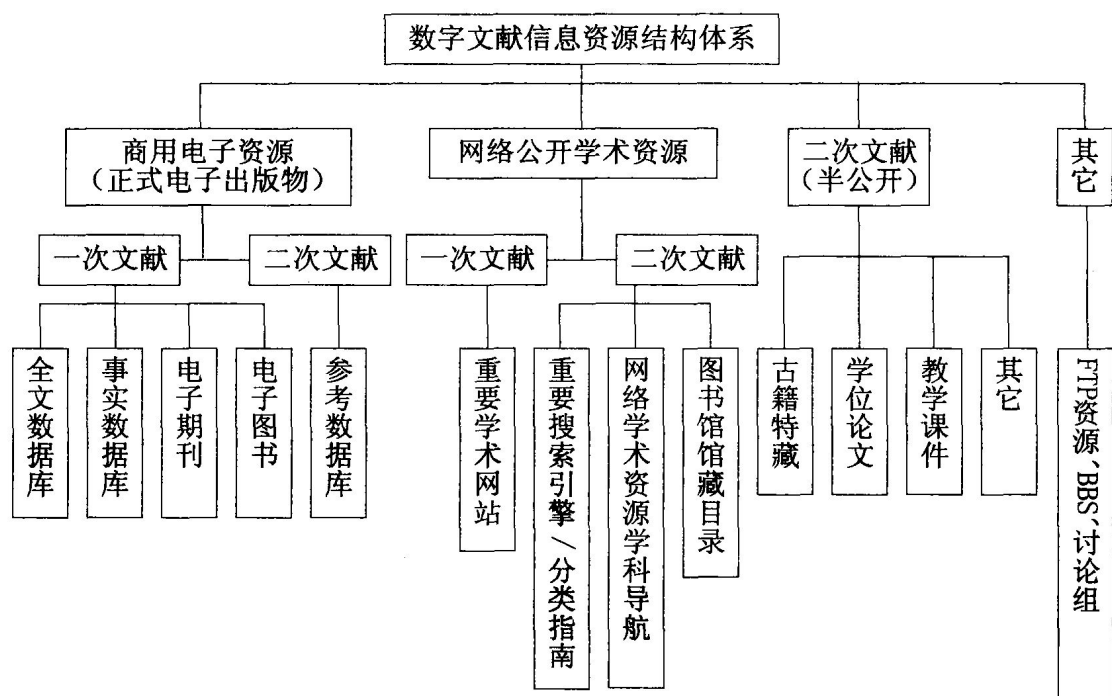


图 1-2-1 数字文献信息资源结构图

商用电子文献资源：也可称电子出版物，是由正式出版机构或出版商 / 数据库商出版发行的，在数字学术信息资源中所占比例最大，包括各类数据库和电子期刊、电子图书。其特点是：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。

网络公开学术资源：属半正式出版物，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、教育机构等在网页上正式发布的网页及其信息，亦属于一次文献类型。使用这部分信息主要依靠搜索引擎 / 分类指南、网络学术资源学科导航等二次文献资源。用于提供使用图书馆印刷型馆藏的联机公共目录（OPAC）也属于这部分范畴。

特色资源：亦属半正式出版物，主要基于各教育机构、政府机关、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，不完全向公众发行，有时需要特别申请。例如教师的教学课件（CAI），只在校园网内的教学范畴允许使用。

其它资源，如 FTP 资源、新闻组、BBS、电子邮件等，属于非正式出版物。

2. 数字型文献的类型。

（1）全文数据库（full-text database），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、商业信息等为主。如清华同方的《中国期刊全文数据库》、美国的《ProQuest Medical Library》期刊全文数据库等。

（2）事实数据库（factual database），指包含大量数据、事实，直接提供原始资料的数据库，又分为数值数据库（numeric database）、指南数据库（directory database）、术语数据库（terminological database）等，相当于印刷型文献中的字典、辞典、手册、年鉴、百科全书、机构指

南、人名录、公式与数表、图册(集)等。数值数据库,指专门以数值方式表示数据,如统计数据库、化学反应数据库等;指南数据库,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、词语信息等数据库,如电子版百科全书、网络词典等。

全文数据库与事实数据库在二十世纪七十年代曾被合称为源数据库(source database),意指文献信息来源的数据库。后随着这两种类型的数据库的发展而逐渐分离。

(3) 参考数据库(reference database),指包含各种数据、信息或知识的原始来源和属性的数据库。数据库中的记录是通过对数据、信息或知识的再加工和过滤,如编目、索引、摘要、分类等,然后形成的。到目前为止,参考数据库主要是针对出版物开发的,目的是指引用户能够快速、全面地鉴别和找到相关的信息。

参考数据库主要包括:书目数据库、文摘数据库、索引数据库。书目数据库主要针对图书进行内容的报道与揭示,如各图书馆的馆藏机读目录数据库;文摘和索引数据库则只对期刊论文、会议论文、专利文献、学位论文等进行内容和属性的认识与加工,如《中国生物医学文献数据库》(CBMDisc)、美国《MEDLINE》、荷兰《EMBASE》等。

(4) 电子图书(eBook),最初的电子图书主要是百科全书、字典词典等工具书,现在已经涉及到了很多学科领域,文学作品和学术专著所占的比例越来越大。

(5) 电子期刊(electronic journals,简称 e-journal),包括:与纸本期刊并行发行的电子期刊,如著名的《科学》杂志(Science)、《自然》杂志(Nature)、中国电子期刊杂志社的期刊等;纯电子期刊,如《数字图书馆杂志》(D-Lib Magazine)。

(6) 电子报纸(electronic newspaper),目前网上已有数千种报纸供用户浏览阅读。同电子期刊一样,电子报纸同样也有印刷型报纸的电子版和纯电子报纸两种类型。

(7) 搜索引擎/分类指南(search engine),是目前利用互联网开放信息的常用工具,也可以称得上是互联网开放信息的索引目录。搜索引擎主要是使用一种计算机自动搜索软件在互联网上检索,将检索到的网页编入数据库中,并进行一定程度的自动标引,用户使用时输入检索词,搜索引擎将其与数据库中的信息匹配,然后产生检索结果。分类指南是将搜索到的网页按主题内容组织成等级结构(主题树),用户按照这个目录逐层深入,直至找到所需的文献。通常搜索引擎与分类指南结合在一起,如Yahoo、新浪等。

(8) 网络学术资源学科导航:将互联网上的开放信息加以甄别、筛选和科学管理,按学科组织起来,构成完整的学科导航系统,为教学、科研、技术人员提供各类学术信息。与搜索引擎/分类指南不同的是,网络学术资源的学科导航库通常是由图书馆单独或联合建设。

(9) FTP资源:FTP含义是File Transfer Protocol,意为文件传送协议,是互联网上最早应用的协议之一,它可以使用户远程登录到远端计算机上,把其中的文件传回到自己的计算机上,或把自己计算机上的文件上传到远端计算机系统上。所谓FTP资源,是指互联网上的开放FTP站点,这些站点允许用户登录上去,从中下载各类数据、资料、软件等。有些搜索引擎,如天网(<http://e.pku.edu.cn>),可以专门用来检索网上的FTP站点。

(10) 其它:如网站、BBS、新闻组等,也可以给用户提供一些有用的知识或动态信息。

三、现代医学文献信息的发展特点

1. 信息的生产量惊人,医学文献数量占各学科之首

信息的生产和发布历来受官方或正式出版机构管制,自从有了互联网后,任何人都可以成为网上信息的无拘束发布者,具有很大的任意性和自由性。于是,信息的生产量随之而持续剧增。人们常说的“互联网年”为3~4个月,即互联网上每3~4个月增长的信息相当于传统方式的一年的信息量,这种发展无疑将超过传统传媒工具——报刊、广播和电视。

医学及生物科学文献的增长总是随人类信息总量的增长而增长。据统计,占人类文献总量70%属期刊文献,而期刊文献中,约1/7是生物医学期刊文献。《science Citation Index》按引文数量排列的前500种期刊中,生物医学期刊占176种,超过了1/3。在其它各类文献中,医学类文献所占比例基本相似。

2. 文献信息的传播速度加快,半衰期缩短

文献信息的传播方式随互联网的兴起而发生了极大的变化。传统的信息传播方式是通过书刊、广播、电视向大众单向传播,互联网则是通过光纤网进行,这使得信息与大众的关系发生了变化。信息与大众之间的关系从单向灌输转为平等交流,大众可以自由选择信息。尽管文献信息没有集中存放,而是分布在全球各地,但由于通过光纤网络传递,过去需要几天甚至几十天才能获取的信息现在在几分钟之内就可以得到,时间和空间的阻隔几乎不再存在。文献信息的传播速度加快,催化了文献的半衰期缩短。在十九世纪,文献的半衰期为50年左右,现在已经缩短到了3~8年,有些学科已经缩短到了以月计算,如信息技术、微电子技术、分子生物学等。

3. 文献信息的分布结构发生变化

过去对专业文献的分布状况进行分析得出结论,约1/3的论文发表在本专业期刊上,1/3的论文发表相关专业期刊上,另有1/3的论文则发表在与本专业无关的期刊上。现在,互联网上有上100亿个网页,发布了无数的一次文献,绝大多数的网页没有学科专业的属性。这种发展趋势无疑是对几十年文献分布规律的冲击。

4. 文献信息的结构发生变化

文献信息的结构发生变化始于上世纪的八十年代,九十年代开始加速。在内容上,由于学术团体、政府机关、商业部门、个人、民间组织等任何组织或个人都可以在网上发布信息,因而对文献信息失去控管,没有认证和审核,使得各种信息甚至有害信息混在一起。在形式上,传统的印刷型图书期刊正在逐渐变为数字化信息,不再仅仅是视觉和静态形式,而是多媒体和动态的,需要强大的计算机软件系统来进行管理和使用。在类型上,不再仅仅是传统的正式出版物,而是电子期刊/图书、非正式出版物、半公开出版的灰色文献、数据库、软件、新闻组、电子公告板、FTP等各类资源共同构成网络环境下的信息资源。

5. 文献的获取与使用发生变化

过去获取文献主要依赖图书馆,现在变成了主要依赖网络。通过网络获取文献信息可以不受任何时间和空间的限制,可以随时随地从网上获取。在文献信息的使用上,可以反复多次地不受任何限制地使用,还可以轻而易举地对一次文献进行二次、三次深度加工。