

医学信息检索

Medical Information Retrieval

方习国 主编



北京师范大学出版集团
BEIJING NORMAL UNIVERSITY PUBLISHING GROUP
安徽大学出版社

医学信息检索

主 编 方习国

副主编 吴义苗

编 委 方习国 胡笑梅 江 婧 李世莉
孟庆杰 孙 俐 吴义苗 肖燕秋
杨 敏 朱 玲



北京师范大学出版集团
BEIJING NORMAL UNIVERSITY PUBLISHING GROUP
安徽大学出版社

图书在版编目(CIP)数据

医学信息检索/方习国主编. —合肥:安徽大学出版社, 2012. 1

ISBN 978-7-5664-0354-4

I. ①医… II. ①方… III. ①医药学—情报检索 IV. ①G252.7

中国版本图书馆 CIP 数据核字(2011)第 271903 号

医学信息检索

方习国 主编

出版发行: 北京师范大学出版集团
安徽大学出版社
(安徽省合肥市肥西路 3 号 邮编 230039)
www.bnupg.com.cn
www.ahupress.com.cn

经 销: 全国新华书店
印 刷: 安徽江淮印务有限责任公司
开 本: 184mm×260mm
印 张: 16.5
字 数: 388 千字
版 次: 2012 年 2 月第 1 版
印 次: 2012 年 2 月第 1 次印刷
定 价: 30.00 元

ISBN 978-7-5664-0354-4

责任编辑: 钟 蕾

装帧设计: 李 军

责任印制: 赵明炎

版权所有 侵权必究

反盗版、侵权举报电话: 0551—5106311

外埠邮购电话: 0551—5107716

本书如有印装质量问题, 请与印制管理部联系调换。

印制管理部电话: 0551—5106311

前 言

本书分为三个部分。第一部分为该书第一章,主要内容为信息检索的基本知识,包括检索系统、检索语言、检索技术等,为了能够引起学生对信息检索的重视,在该部分还专门增加了大学生信息素养的内容。第二部分为该书第二章到第八章,具体介绍各种各类与医学相关的信息资源,重点是其检索方法。包括:数字图书馆、医学文摘数据库、医学全文数据库、医学网络信息资源、引文数据库;此外还包括:专利等特殊类型信息资源、循证医学和生物信息学等学科专题信息资源。第三部分为该书第九章和第十章,内容涉及检索效果的评价和分析,以及论文写作等。

参加该书写作的主要是在图书馆工作且从事医学文献检索的一线教师。在该书内容的选择上,既介绍最新的检索知识和相关数据库,又尽量贴近实际工作和教学的需要。

本书在写作的过程中,参考了大量的专家与学者的研究成果,在此一并表示感谢!

本书的出版得到蚌埠医学院教材科李玉良同志的鼓励和支持,特表示感谢!

由于时间紧迫和编者水平有限,本书难免会存在一些错误,恳请大家批评指正!

编 者

2011 年 12 月

目 录

第一章 信息检索基础	(1)
第一节 信息概述	(1)
第二节 信息检索系统	(7)
第三节 信息检索语言	(11)
第四节 文献分类法	(14)
第五节 信息检索技术、途径和步骤	(18)
第六节 大学生信息素养	(22)
第二章 文摘数据库	(27)
第一节 中国生物医学文献数据库(CBM)	(27)
第二节 PubMed	(34)
第三节 EMBASE 数据库	(57)
第四节 BIOSIS Previews	(65)
第五节 SciFinder Web	(69)
第三章 全文数据库	(76)
第一节 中国学术期刊网络出版总库	(76)
第二节 维普资讯	(82)
第三节 万方医学网	(90)
第四节 EBSCO 全文数据库	(96)
第五节 Elsevier 全文数据库	(100)
第六节 SpringerLink	(104)
第四章 数字图书馆	(107)
第一节 超星数字图书馆	(108)
第二节 方正 Apabi 数字资源平台	(113)
第三节 书生之家数字图书馆	(115)
第四节 NetLibrary 电子图书	(117)

第五章 引文数据库	(121)
第一节 引文检索概述.....	(121)
第二节 Web of Science	(124)
第三节 中国科学引文数据库.....	(132)
第四节 其他引文检索资源.....	(137)
第六章 特殊类型信息资源	(145)
第一节 专利信息检索.....	(145)
第二节 学位论文信息资源.....	(154)
第三节 学术会议信息资源.....	(161)
第四节 标准信息检索.....	(170)
第七章 网络信息资源	(176)
第一节 网络信息资源概述.....	(176)
第二节 搜索引擎.....	(178)
第三节 医学搜索引擎与医学网站.....	(182)
第八章 学科专题信息资源	(190)
第一节 循证医学信息资源.....	(190)
第二节 生物信息学资源.....	(196)
第三节 药学信息资源.....	(202)
第九章 医学信息检索应用与评价	(211)
第一节 信息检索效果评价.....	(211)
第二节 医学文献检索策略与案例分析.....	(213)
第三节 医学信息分析.....	(221)
第四节 网络医学信息资源评价.....	(230)
第十章 医学科研论文写作	(235)
第一节 医学科研信息查新.....	(235)
第二节 医学科研论文的基本结构与要求.....	(242)
第三节 医学科研论文的写作步骤与特点.....	(251)
参考文献	(255)

第一章 信息检索基础

科学技术的迅猛发展使知识的数量急剧增长,知识的老化速度也明显加快,这就要求人们具备终身学习的能力。这种能力在很大程度上就是获取新信息和新知识的能力,就是对新知识的敏感力和接受力,因此必须掌握信息检索的方法。在信息化社会,获取信息、分析评价信息、吸收、整合和利用信息能力更显重要。信息检索是获取知识的捷径,因此掌握信息检索方法与技能,可以帮助人们快速、高效地获取所需信息,并对信息进行分析、评价和有效利用。

第一节 信息概述

一、信息、知识、情报和文献的概念

1. 信息

在现代信息社会中,物质、能量与信息已成为人类生存和发展的三大要素。信息与人类的生产和生活息息相关,我们通过信息认识物质、认识能量、认识周围世界。

信息至今还没有一个公认的定义。国家标准《情报与文献工作词汇基本术语》(GB/T 4894-2009)中信息的定义为:“信息是物质存在的一种方式、形态或运动状态,是事物的一种普遍属性,一般指数据、消息中包含的意义,可以使消息中所描述事件的不确定性减少。”因为信息是事物的一种普遍属性,是人对客观事物属性以及运动状态的感知,所以信息也就有了自然信息和社会信息之分。自然信息可表现自然界中事物的状态及特征等,例如春、夏、秋、冬等自然现象;而社会信息反映人类社会各种事物和现象的状态及特征,例如人的喜、怒、哀、乐等。在科学技术领域,人们通过科学研究获得的或与科学研究活动相关的信息称作科技信息。医学信息是指通过观察、实验或借助于其他工具,对健康/疾病人体生理或病理状态特征的认识及其反映。例如,人体脉搏、呼吸、体温以及疾病状态下的各种体征与症状、实验室及医技科室的检测数据、医学图像等都是医学信息。

信息被认为是无所不有地存在于自然界、人类社会及思维领域中,所以信息的产生、传递、接收是自然界和人类社会一种极为普遍的现象。人类社会发展的历史就是人类不断认识信息、获取信息、掌握信息、传递信息、生产信息,并用其为人类服务、改造客观世界也改造主观世界的过程。信息区别于其他事物的本质属性,主要表现在以下几个方面:

(1) 可识别性。信息作为客观事物的反映,能够通过人的感受被接收与识别。识别

又可分为直接识别和间接识别,直接识别是指通过感官的识别,间接识别是指通过各种测试手段的识别。不同的信息源有不同的识别方法。

(2) 可扩充性。信息随着时间的变化,将不断扩充。信息永远都在产生、更新、演变,是取之不尽、用之不竭的智慧源泉。

(3) 可压缩性。人们对信息进行加工、整理、概括、归纳,使之精练且浓缩,便于利用。

(4) 可传递性。信息是可以通过各种方法存储的,即经人类感知、捕捉、获取的信息,可以借助于各种载体或媒介广泛传播,为更多的人所接收和利用。

2. 知识

知识是人类的主观世界对客观世界的概括和如实的反映,根据所获得的信息,上升为知识。知识是人类对事物本质的认识和经验的总结归纳,也是被认为正确真实,可以指导解决实践问题的观点、经验等信息。知识来源于信息,“是信息的一部分,即被人们理解和认识并以大脑重新组织和系列化的那部分信息”。知识是人类的宝贵财富,是人类的智慧结晶。它具备以下多种附加特征。

(1) 隐性特征。知识具备较强的隐蔽性,需要进行归纳、总结、提炼。因为这些隐性知识是自己的个人经验、对事情的感悟和深层次的理解等方面长期地积累而形成的知识。隐性知识显性化就是知识分享的过程。

(2) 行动导向特征。知识能够直接推动人的决策和行为,加速行动过程。

(3) 动态特征。知识在应用、交流的过程中,被不断丰富和拓展,不断更新和修正。

(4) 主观特征。每个人对知识的理解都会加入自己的主观意愿。

(5) 可复制/转移。知识可以被复制和转移,可重复利用。

由此可见,知识是一部分系统化、理论化的信息。如医学知识是人们通过实践对医学信息的获取、提炼和系统化、理论化的结果,是关于人体生活、健康、疾病的现象、本质和规律的认识。知识大致可分为两个部分,即显性知识和隐性知识。显性知识是指记录在各种介质上的知识,如图书、档案、数据库等,属静态知识;隐性知识是指存在于人头脑中的未被编码的经验性知识,如个人的技术诀窍、想象与创意等,属于活跃性的动态知识。两者可相互转化。知识按多种标准可划分为不同类型,如自然科学知识、社会科学知识、哲学知识、思维科学知识等。

3. 情报

关于情报的定义,学术界还没有定论。中国早期对情报的定义是:“战时关于敌情之报告。”后来,情报概念有了新发展,认为情报是“针对一定对象的需要而传递有参考价值的新信息和新知识”;“情报就是为解决(或针对)某个(些)特定问题的要求,从传递交流的信息中提取的适用知识”等。换言之,情报就是为了解决一个特定问题所要的知识,是激活了、活化了的,以及这种知识的及时性和针对性。

情报是一种普遍存在的社会现象,人们在物质生产和知识生产的实践活动中,源源不断地创造、交流与利用各种各样的情报。情报是一种针对具体需要的适用知识,适用是它的终极目的。无论情报的内容与形式如何变化,其不变之处是具有三个基本属性:知识性、传递性和效用性。

(1) 知识性,是指以知识为实体。情报的本质就是知识,情报都包含有知识或信息,所以知识和信息是构成情报的原料;但并非所有的知识和信息都能构成情报,只有经过筛

选、加工,为用户所需的新知识或新信息才成为情报。知识性是情报最主要的属性。

(2) 传递性,是指以传递为手段。人们为一定目的搜集的有使用价值的知识或信息,还必须经过交流传递,并被用户接受或利用,才能成为情报。知识或信息若不进行传递交流、供人们利用,就不能构成情报。传递性是情报的第二基本属性。

(3) 效用性,是指以效用为目的。情报以实现其使用价值为目的,人们创造情报、传递情报的目的在于利用,在于提高其效用性,效益是情报的结果。只有信息源产生的信息被人们以某种方式接收,在一定的时间内经过用户使用并产生效益,才能成为情报。因此,情报为特定用户服务,用户需要各自情报。时间上的效用性是衡量情报好坏的重要标志。

4. 文献

国家标准《文献著录总则》(GB3792. 1-83) 关于“文献”的定义是:“文献:记录有知识的一切载体。”在这一定义中,有两个关键词:“知识”是文献的核心内容,“载体”是知识赖以保存的附着物。总而言之,文献是指以文字、图像、符号、声频、视频、编码等手段将信息、知识记录或描述在一定的物质载体上,并能起到存储和传播信息、情报和知识的一切载体。文献是记录有知识的载体,由三个基本要素构成:

(1) 内容是关键,即内容上要有知识或信息。

(2) 符号是表现形式,即具备揭示和表达知识信息的标志符号,如文字、图像、声频、视频等。

(3) 载体为文献存在的方式,即记录信息符号的一切物质载体,如纸、光盘、磁盘等。

文献是人们获取知识的重要媒介,文献的内容反映了各个时代、各种社会环境下科学和人类社会进步所达到的知识水平状况,记录了人类历史长河中科学技术发展与人类活动中所达到的成就和水平,凝结着人类的辛勤劳动和智慧,积累着各种对后人有用的事实、数据理论、方法。人们又从文献中汲取、利用知识贡献于社会,从而极大地推动了社会文明的发展。

人类认识社会与自然界的各种知识的积累、总结、贮存与提高,主要是通过文献的记录、整理、传播和研究而实现的。文献能使人类的知识超越时空地保存和传递。文献也是科学研究的基础。任何一项科学研究都必须广泛搜集文献资料,在充分占有资料的基础上分析资料的种种形态,探求其内在的联系,进而作更深入的研究,获取更多的科研成果。文献的基本功能有存储知识信息、传递知识信息、教育和娱乐功能等。

二、信息、知识、情报和文献的关系

信息、知识、情报和文献之间有着极为密切的关系。信息、知识、情报必须固定在一定的物质载体上,形成文献后才能进行传递,被人们所利用。文献是信息、知识、情报存储、传递、利用的重要方式。信息可以成为情报,但是一般要经过选择、综合、研究、分析等一系列加工过程;信息是知识的重要理论组成部分,但不是全部,只有系统化、理论化的信息才能称作知识;情报是知识或信息经传递起作用并产生效益的那部分。简言之,信息是人对客观事物属性以及运动状态的感知;知识是对信息认知的那部分内容,是信息的一部分;知识只有通过特定的时间传递给特定的用户才能成为情报;文献是知识的一种载体。

由上述可见,四者是包含与被包含的关系,知识是信息中的一部分,情报是知识中的一部分,文献又是知识的一种载体。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、情报和文献的相容关系如图 1-1 所示。

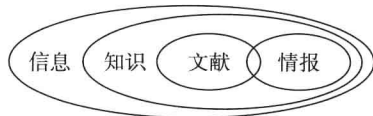


图 1-1 信息、知识、情报和文献的相容关系

三、信息资源的类型

信息资源,是指以各种记录形式存在的信息载体的总和。信息是多种类、多层次、多方面的。信息的类型可根据不同的划分标准,区分为不同的类型,按载体的不同可分为实物信息、口头信息、文献信息和网络信息四大类。常将文献信息划分为如下几种类型。

1. 按载体形式划分

(1)印刷型。印刷型文献主要是以纸张为信息载体、以印刷技术为记录手段而形成的纸质出版物,如图书、期刊等。它的优点是便于阅读,符合传统阅读习惯并可广泛流传;缺点是信息存储密度小、不宜长期保存、占空间多,整理、存储均需花费较多人力和物力。

(2)缩微型。缩微型文献是以感光材料为存储载体,利用光学记录技术,把文献影像缩小并记录在胶卷或胶片上的文献,如缩微胶卷、缩微胶片和缩微照片等。它的优点是体积小、存储密度高、便于复制;缺点是必须借助显微阅读器才能阅读。

(3)视听型。视听型文献又称“声像资料”,指利用声像技术直接记录声音、图像所产生的一种文献形式,然后通过播放手段给人以听觉、视觉感受的文献。如唱片、录音带、电影拷贝、录像带等。这种文献直接记录声音和图像,其生动的动态信息让人们更为直观地感受和认识事物。

(4)电子型。电子型文献是采用电子手段将信息和知识数字化储存于光盘、磁盘等载体上,或通过通信网络传送到用户终端供人们利用的出版物。它是通过编码和程序设计把文献变成数字语言和机器语言输入到计算机中去,存储在磁带、磁盘或磁鼓上,需要时由计算机输出。电子型文献是一种必须借助于计算机及现代化通讯手段传播利用的一种新的文献类型,如各种类型的数据库、电子期刊、电子图书等。

电子出版物的问世是信息时代的重要标志,它改变了书刊的物理形态,冲击了人们的传统阅读习惯,开辟了一种全新的信息传播渠道,极大地提高了信息的传递效率,加快了社会信息化的进程。

2. 按出版形式划分

(1)图书。图书是现代出版物中最普通的一种类型,内容比较成熟、定型,是系统学习和掌握各门科学知识的最重要资料。图书种类繁多、形式多样、功能各异,对学习与研究而言,常用的图书主要有教科书、专著、参考工具书等。

每一种正式出版的图书在版权页或其他明显部位都标有一个由 10 位或 13 位(2007 年 1 月 1 日 ISBN 的格式由 10 位修订为 13 位)数字组成的国际标准书号(International

Standard Book Number,简称 ISBN)。ISBN 是一种国际通用的出版物代码,代表某种特定图书的某一版本,具有唯一性和专指性,读者可依据该号码通过某些文献信息系统查询该种图书。

(2)期刊。期刊又称杂志,是一种连续出版发行的文献,期刊有固定的名称(刊名)、相对固定的版式、一定的出版规律。定期并长期连续出版,是连续出版物的主要类型。每期发表多篇文章,作者众多,内容新颖,出版周期短,信息量大,约占整个信息量的 70%,是情报的主要信息源,且反映的多数是最新的科技成果。

期刊按内容和用途可分为理论性或学术性期刊、技术性期刊、知识普及性期刊、检索性期刊等;按出版规律可分为定期和不定期两种;按出版频率可分为月刊、半月刊、周刊等。

每一种经申请获准出版的连续出版物都可得到一个固定不变的国际标准刊号。ISSN 是国际上统一用于识别连续出版物的标准编码系统,具有唯一性和专指性,可以成为读者查询某种刊物的一个检索途径。

(3)会议文献。会议文献指在学术会议上进行交流的论文,会后汇编成册,如会议论文集、会议论文汇编、会议辑要等。会议文献一般是最新研究成果或阶段性成果,内容比较新颖,能把握科研动态,学术性强,是重要的情报信息来源。

(4)学位论文。学位论文是指在校的学生为获得学位而撰写的论文。学位论文国家标准(GB7713-87)将其定义为:“学位论文是表明作者从事科学研究取得创造性的结果或有了新的见解,并以此为内容撰写而成,作为提出申请授予相应的学位时评审用的学术论文。”学位论文包括学士论文、硕士论文、博士论文。学位论文的特点是探讨问题专一,论述详细系统,数据充分,有新论点、新依据,具有一定的独创性,对科研和教学有一定的参考价值。目前有相当多的数据库收集学位论文,如中国学术期刊网、万方数据库等。

(5)专利文献。专利文献指专利局公布或归档的与专利有关的文献,包括专利说明书、专利公报、专利分类资料等。专利是指受到法律保护的技术发明,是知识产权的一种具体体现形式。专利文献是各国及国际性专利组织在审批专利过程中形成并定期出版的各类文献的总称。专利说明书是专利发明人或申请人为获得某项发明专利权向国家专利主管部门递交的有关该发明创造的详细技术说明书,内容包括发明的目的、用途、特点、效果、工艺过程、技术细节,并附图表和各种数据。专利文献反映了当时某项科技所达到的最新成就,属于特殊类型的科技文献,也是科学技术领域内的一种重要信息来源。

(6)科技报告。国家标准(GB7713-87)给出的定义为:“科技报告是描述一项科学技术研究的结果或进展或一项技术研制实验和评价的结果,或是论述一项科学技术问题的现状和发展的文献。”科技报告是科技人员对某学科或课题进行研究的阶段报告、成果报告和总结报告,涉及的内容广泛、专深具体而且是最最新的研究成果,往往能反映出一个国家或一个专业的科研水平。

3. 按信息加工深度划分

(1)一次文献。即原始文献,是指作者以其本人的研究成果(如观察、实验、调查研究等的结果)为基本素材写成的原始创作,包括期刊论文、专著、科技报告、专利说明、会议论文、学术论文、技术标准等。

一次文献能反映科学技术发展的前沿,内容具有一定的创新性。一次文献所记录的

是作者的最新发现、新的见解、新的理论、新的方法、具体而详尽的知识,是人们学习参考的最基本的文献类型,也是最主要的文献情报源。一次文献是产生二次、三次文献的基础,是文献检索的主要对象。

(2)二次文献。二次文献通常称为检索工具,是对一定范围、时间或类型的大量无序、分散的一次文献进行收集、整理、压缩、加工,著录其特征如篇名、著者、分类、主题、出处等,并按一定的顺序组织编排,形成供读者检索所需一次文献线索的新的文献形式,如书目、索引、文摘、题录及相应的书目数据库。二次文献因其具有检索功能而被称为检索工具或检索系统,帮助读者在较少的时间内获得较多的文献信息。二次文献的重要性在于能够为查找一次文献提供线索,通过二次文献查找大量分散的一次文献。

(3)三次文献。三次文献是对一次和二次文献进行系统整理、综合分析后编写而成的概括性文献,如综述、评论、进展、年鉴、手册、指南等。三次文献的特点是:信息含量大、综合性强、参考价值大,读者不必大量阅读一次文献就可借此比较全面地了解某一专题或某一领域的研究水平及动态。另外,它也是读者掌握情报源的主要资料。

(4)零次文献。零次文献指未经信息加工,直接记录在载体上的原始信息,即未正式发表或未进入社会交流的最原始的文献,如观测记录、实验数据、调查材料、设计草图等。它是一次文献的素材,对一次文献的形成具有重要作用。

综上所述,各级别文献的形成及其关系可用图 1-2 表示。

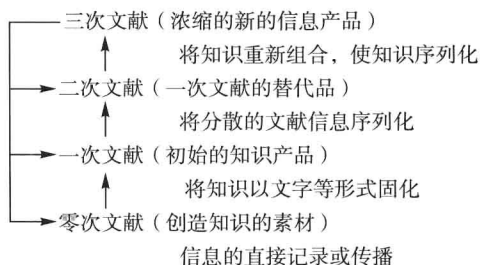


图 1-2 各级别文献的形成及其关系

图 1-2 不仅展示了各级别文献的递进关系,更揭示了知识产生及有序化的进程。这一综合的系统过程的不断循环往复,既是一种文献信息工作,更是以知识信息的开发利用为前提的知识创新活动。

四、信息资源的特点

随着科学技术的迅猛发展与人类社会文明进程的日益加快,文献信息亦呈现出相应的变化,主要表现在以下几个方面。

1. 数量多、增长快

科学技术的迅猛发展造成科技文献数量急剧增长。当前,全世界的期刊数量已超过 100 万种,来自 200 多个国家和地区,每年发表的论文数量超过 400 万篇。我国的科技期刊已超过 4800 种,年发表论文超过 20 万篇。从整个世界科技出版物来看,医学文献占整个科技文献的四分之一,例如常用的外文生物医学期刊有 3000 多种,国内出版的生物医学期刊达 1700 多种。由此可见,医学文献增长十分迅速,数量非常庞大,占各学科之首。

2. 出版类型复杂、文种繁多

目前科技文献的载体除了传统的印刷型外,还有各种视听型、电子型文献,载体的多样化,大大方便了读者。在相当长的时期内,印刷型文献将与其他类型文献同时并存,互相补充。随着科学的发展,文献出版的语言种类急剧增多,如美国《化学文摘》收录的语种有 50 多种,美国《医学索引》收录的语种有 40 多种。文种的增加丰富了科技文献,拓宽了研究空间,同时也增加了读者阅读各种文献的障碍。

3. 文献分布异常分散,内容重复交叉

由于受多种因素的影响,文献分布异常分散,内容重复交叉发表的现象呈上升趋势。就医学领域而言,由于学科之间相互交叉、渗透,新型学科、边缘学科层出不穷,使得医学文献不仅刊载在医学专业期刊上,还大量刊载在一些综合性期刊或其他相关学科领域的期刊上。现代科学技术综合交叉、彼此渗透,一种专业刊物上报道的内容往往包含了多个学科,一篇专题论文又往往涉及几个专业,且有同一篇文章用不同形式、不同文字,在不同范围内多次发表的现象。会议论文、学位论文、科技报告除了以自己独特的形式出版外,经过一段时间后又会上转化为其他形式出版。这种交叉重复大大增加了文献量的冗余,也会造成检索上的困难。

4. 知识信息更新加快

科学技术的迅速发展,使得知识更新与信息交流的速度随之加快,导致文献使用寿命缩短。由于文献的出版往往落后于科学技术的发展,有些文献还未出版或刚出版就被新的知识所替代,文献代谢频繁。科技文献的老化周期已由原来的 50 年缩短到现在的 5~7 年。所以当今的人们必须进行继续教育,终生学习新的知识,才能跟上科技发展的速度。

5. 交流传播的速度加快

以往人们以书信、期刊论文等形式互通信息,进行学术交流,现在由于电子计算机、多媒体、通信卫星和因特网技术的广泛应用,信息的存储和传递方式发生了质的飞跃,我们可以通过 WWW、FTP 等方式轻而易举地获得全球信息。Internet 技术的发展,使得信息传递的方式以及人们的阅读方式发生了深刻的变化。

(李世莉)

第二节 信息检索系统

一、信息系统的概念、功能、种类

1. 概念

信息系统是通过计算机、网络等现代技术,从一定的业务领域获得信息,并对信息进行组织、存储、管理、检索和传输等,进而服务于用户的人机结合信息系统。

2. 基本功能

信息系统的基本功能包括:

(1)信息采集。信息采集是信息系统的基础工作。其任务是把相关信息收集起来,转化成信息系统所需的格式数据。信息采集有人工录入数据、网络获取数据、传感器自动采集数据等多种方式和手段。

(2)信息存储。信息存储有物理存储和逻辑存储之分。物理存储是指将信息存储在适当的介质上;逻辑存储是指根据信息的逻辑内在联系和使用方式把大批的信息组织成合理的结构,它的实现常依靠数据存储技术。信息存储要保证数据的一致性、完备性和安全性。

(3)信息处理。信息处理是系统的核心功能。系统的信息经过处理后才能被人们更好地利用。信息的处理包括排序、分类、归并、查询、统计、预测、模拟以及各种运算。例如,校园卡系统中对其系统的数据进行统计、结算、预测分析等都需对数据做数学运算,从而得到管理者所需的相关信息。

(4)信息查询与检索。信息查询与检索使用的是数据库技术和方法,不仅能够实现单项查询,还能够实现组合查询和模糊查询等。

(5)信息传输。信息传输主要包括将采集到的数据传送到处理中心,或从一个子系统传到另一个子系统等。

(6)数据输出。对加工处理后的数据,根据不同的需要,以不同的形式和格式进行输出,例如显示、打印或向其他系统输入等。

3. 种类

信息系统种类繁多,按照不同的划分方法可以分成不同的类型。一般按照它的发展历史和解决的关键问题可以分为:数据处理系统、管理信息系统和决策支持系统。信息系统的应用是随着信息技术、管理理念、组织理念的发展而不断变化的,起初被简单地理解为计算机在数据处理中的应用,后来被理解为管理信息系统,用以辅助企业的管理和决策,因此又被称为“决策支持系统”。

(1)数据处理系统是为了提高事务的处理效率,实现业务的过程监控和对异常情况的反馈,以数据报告为主,解决业务操作人员和一线管理人员的事务问题。例如财务、仓库、销售等业务,主要运用计算机的计算功能。

(2)管理信息系统旨在通过信息共享实现管理的有效性,方便地获得有关业务流程的管理和支持组织的业务信息,解决中高层管理人员在管理中的问题。

(3)决策支持系统更加关注系统外部的信息,能灵活地运用模型与方法对数据进行加工、汇总、分析、预测,得出所需的综合信息与预测信息,解决组织变革中的最大问题,为组织高层提供决策参考信息。

二、信息系统的构成

信息系统由计算机和网络硬件、软件、数据和通讯系统等部分组成,此外人在系统中也起到决定性的作用。

1. 硬件

硬件包括网络服务器、交换机、路由器和光纤等设备。

服务器是在网络上为客户计算机提供各种服务的高性能计算机,作为网络的节点,存储、处理网络上的数据与信息,包括处理器、硬盘、内存、系统总线等部件。其构成与微型计算机相似,但它在处理能力、稳定性、可靠性、安全性、可扩展性、可管理性等方面要优于

微型计算机。

交换机是一种在局域网系统中完成信息交换功能。它基于介质访问控制地址识别, 将其存放在内部地址表中, 通过在数据帧的始发者和目标接收者之间建立临时的交换路径, 使数据帧直接由源地地址到达目的地, 在网络上完成数据包封装和转发等功能。

路由器是连接因特网中各局域网、广域网的设备, 它会根据信道的情况自动选择和设定路由, 以最佳路径、按前后顺序发送信号的设备。路由器是通过路由表来实现路由转换的。通过路由器的一站站转发, 数据包最终沿着某一条路径到达目的地。

光纤是利用光的全反射原理使光在玻璃或塑料制成的纤维中传播, 因光的衰减非常小, 可实现远距离传输。光信号不受外界干扰的影响, 也不会向外界辐射, 这使得光纤传输非常安全。使用光纤时, 要通过某种设备将计算机系统中的电脉冲信号与光脉冲信号进行转换。

存储设备是指计算机的外存储器, 主要包括磁盘机和磁带机。磁盘机以磁盘为存储介质, 具有存储容量大、数据传输率高、容错性能好等特点, 常用于存放大容量的数据。磁带机通常用作数据备份。

信息系统的不同硬件设备组合可以构成多种结构。

(1) 集中式系统结构。整个系统只有一台计算机和相应的外部设备进行数据处理, 适合小型的信息系统。其优点是信息资源集中, 管理方便; 缺点是系统脆弱, 主机的瘫痪可能导致整个系统的崩溃。

(2) 分布式系统结构。利用网络把分布在不同地点的计算机硬件、软件、数据等资源联系在一起, 服务于一个共同的目标, 实现资源共享。其优点是系统扩展方便、稳定性好; 缺点是维护和管理困难。

2. 软件

计算机软件包括程序及其说明和使用指导等。“程序”是指一组指示计算机每一步动作的指令, 是完成一定任务或产生一定结果的指令集合。软件分为系统软件和应用软件。

系统软件负责管理计算机系统中各种独立的硬件, 合理调用计算机资源, 使得它们可以协调工作, 担负着扩充计算机的功能。系统软件使得计算机使用者和其他软件将计算机当作一个整体而不需要顾及到底层每个硬件是如何工作的。系统软件主要包括操作系统、计算机语言系统及数据库管理系统等。计算机操作系统有 Windows 系列、Linux、Unix、Mac OS, 等等; 各种计算机语言编译或解释软件包括 C、C++、VC、VB、VF、Delphi, 等等; 数据库管理系统主要有 Oracle、DB2、Informix、SQL Server、MySQL, 等等。

与系统软件相反, 应用软件是为了某种特定的用途而被开发的软件。不同的应用软件根据用户和所服务的领域不同而提供不同的功能, 可分为通用应用软件和专业应用软件两类。前者如图形图像处理软件、统计分析软件等; 后者如各种专业问题求解软件、人机界面软件等。

3. 数据

在计算机科学中, 数据是对客观事物的性质、状态以及相互关系等进行记载的物理符号或是这些物理符号的组合。各种字母、数字符号的组合、语音、图形、图像等统称为数据。数据来源于客观世界, 经过人脑主观的加工, 最终形成计算机能够处理的信息世界。客观世界存在于人脑之外, 例如一栋房子、一个人、一个组织、一笔交易等。人们必须通过

文字符号、声音或者图像等手段,对这些客观事物的某方面属性进行描述。主观世界是现实世界在人们头脑中的反映,人们通过对其特性的理解和描述来理解客观的存在,例如通过姓名、性别、年龄、籍贯、学历、相貌等来了解一个人。利用数据库技术可组织信息世界,例如通过实体来区分客观存在,通过属性来描述实体。信息世界中的信息在机器世界中的数据存储形式涉及字段、键、记录、文件等。

有组织的数据是信息系统的最基本要素。数据与信息之间既有联系又有区别。数据是记录下来可以被鉴别的符号,信息是对数据的解释。数据经过处理后,经过解释才能成为信息。信息是经过加工后、并对客观世界产生影响的数据。

三、数据库

数据库是存储在一起的相关数据的集合,这些数据是结构化的,无有害或不必要的冗余,并为多种应用服务;数据的存储独立于使用它的程序;对数据库进行插入新数据、修改和检索原有数据等操作均能按一种公用的和可控制的方式进行。当某个系统中存在结构上完全分开的若干个数据库时,则该系统包含一个“数据库集合”(J. Martin 给出的数据库定义)。

数据库的特点有:①实现数据共享;②减少数据的冗余度;③数据具备独立性;④数据实现集中控制;⑤数据具有一致性和可维护性,以确保数据的安全性和可靠性;⑥故障恢复。

数据库的模型有:层次、网状和关系数据库系统。

(1) 层次结构模型:层次结构模型实质上是一种有根结点的定向有序树。有且仅有一个结点无父结点,这个结点称为根结点;其他结点有且仅有一个父结点。按照层次数据结构建立的数据库系统称为层次数据库系统。

(2) 网状结构模型:网状模型是一个网络。它允许一个以上的结点无父结点;一个结点可以有多个父结点。按照网状数据结构建立的数据库系统称为网状数据库系统。

(3) 关系结构模型:关系式数据结构把一些复杂的数据结构归结为简单的二元关系,数据的逻辑结构是一张二维表。由关系数据结构组成的数据库系统称为“关系数据库系统”。每一列中的分量是类型相同的数据;表中的分量是不可再分割的最小数据项;列和行的顺序可以是任意的。在关系数据库中,对数据的操作几乎全部建立在一个或多个关系表格上,通过对这些关系表格的分类、合并、连接或选取等运算来实现数据的管理。元组是关系集合中的一个元素,是表中的一行,或称一条记录。属性是关系中每个域的含义,是表中的一列。表的列也称为“字段”,每个字段包含某一专题的信息。例如,信息检索数据库中,“作者”、“文献类型”等就是字段,它们处于表的同一列中,属性相同。域是属性的取值范围。分量是元组中的属性值,是某行某列的值。码是某个属性的值能够唯一标志关系的元组,能够唯一标志二维表中的一行。目前占据主导地位的数据库管理系统有 Oracle, Sybase, Informix, Visual FoxPro, SQL Server, MySQL 等。

索引是对数据库表中一列或多列的值进行排序的一种结构。建立索引的目的是加快对表中记录的查找或排序。

四、信息检索系统

信息检索系统是指为满足信息用户的信息需求而建立的,运用特定的信息收集、管理

域检索等技术,存贮经过加工的信息集合,提供一定存贮方法与检索服务功能的一种相对独立的服务实体。医学上常见的信息检索系统有文摘型数据库检索系统、全文型数据库检索系统、网络搜索引擎等。

一个完整的信息检索系统,通常由信息源、信息组织管理、用户接口和提问处理等几个有机部分组成。

1. 信息源

信息源是指计算机检索系统信息或数据的来源。信息检索系统中的数据主要来自各种公开文献,如一次文献中的期刊、图书、研究报告、会议论文、专科文献、政府出版物、学位论文,二次文献中的摘要、索引和目录,三次文献中的百科全书、专科词典、名录、指南、手册等。还有网络上的公开信息。每个系统根据其目标和服务对象的需要,确定数据收集范围,并广泛地采集各种信息源,为系统提供充足而适用的数据。

2. 信息组织管理

信息组织管理主要是指信息标引的方法、组织方式和更新周期。信息组织管理科学、实用、合理与否,直接关系到信息检索的效果好坏。标引就是根据系统的规则和程序,对文献内容进行分析,然后赋予每篇文献一定数量的内容标志,如分类号、主题词、关键词等,作为存储与检索的依据。标引的主要依据是词表。主题词表是控制标引用词和检索用词,使二者尽可能取得一致的有效工具。信息组织中一项重要的工作是建立索引,它能够提高系统检索效率。

3. 用户接口

用户接口是面向系统用户的人—机接口程序,它承担着用户与系统之间的交流功能。用户接口提倡界面友好、形象与直观,还考虑使用时的简便性和反馈机制等。接口部分使用的命令语言包括基本命令,如数据库选择、选词、组配、结果输出、求助等,以及功能扩展命令,如截词、位置运算、限制检索、保存检索策略等。

4. 提问处理

用户输入检索词或提问式后,系统要将检索词或提问式与数据库中存储的数据进行比较运算,然后把运算结果输出给用户。检索程序操作过程如下:①接收提问;②校验提问,包括语法检查、格式检查和用词检查;③加工提问,指对源提问式进行解释性或编译性的加工,生成便于机器处理的目标提问式;④检索,即从数据库中读入一批记录,与提问式进行比较,把满足要求的记录送入输出文档。

(方习国 孟庆杰)

第三节 信息检索语言

一、信息检索语言的概念

信息检索语言是组织文献与检索文献时所用的语言,是一种专门供信息标引和信息检索使用的人工语言,又称索引语言、标引语言等,是信息检索领域中用来描述信息内容