



“十二五”普通高等教育本科国家级规划教材



国家卫生和计划生育委员会“十二五”规划教材

全国高等医药教材建设研究会“十二五”规划教材

全 国 高 等 学 校 教 材

供8年制及7年制（“5+3”一体化）临床医学等专业用

医学文献信息检索

Medical Literature Information Retrieval

第3版

主 编 罗爱静 于双成

副主编 马 路 王虹菲 周晓政

Medical science
foundation

Medical professional
attitude, behavior and ethics

Clinical skills

Information management capacity

Critical thinking

Group health and health system

Communication skills



人民卫生出版社
PEOPLE'S MEDICAL PUBLISHING HOUSE





“十二五”普通高等教育本科国家级规划教材

国家卫生和计划生育委员会“十二五”规划教材
全国高等医药教材建设研究会“十二五”规划教材
全国高等学校教材

供8年制及7年制(“5+3”一体化)临床医学等专业用

医学文献信息检索

Medical Literature Information Retrieval

第3版

主 编 罗爱静 于双成

副主编 马 路 王虹菲 周晓政

编 者 (以姓氏笔画为序)

于双成 (吉林大学公共卫生学院)

马 路 (首都医科大学)

王虹菲 (大连医科大学)

仇晓春 (上海交通大学)

孔瑞珍 (山西医科大学)

史继红 (哈尔滨医科大学)

任慧玲 (北京协和医科大学)

宋余庆 (江苏大学)

张士靖 (华中科技大学同济医学院)

罗爱静 (中南大学湘雅三医院)

周旭毓 (中山大学)

周晓政 (南京医科大学)

赵文龙 (重庆大学)

胡德华 (中南大学湘雅医学院)

崔 雷 (中国医科大学)

符礼平 (复旦大学)

谢志耘 (北京大学)

学术秘书 戴萍萍 (中南大学湘雅三医院)

人民卫生出版社

图书在版编目(CIP)数据

医学文献信息检索 / 罗爱静, 于双成主编. —3 版. —北京:
人民卫生出版社, 2015

ISBN 978-7-117-20571-9

I. ①医… II. ①罗…②于… III. ①医药学—情报检
索—医学院校—教材 IV. ①G252.7

中国版本图书馆 CIP 数据核字(2015)第 070555 号

人卫社官网	www.pmph.com	出版物查询, 在线购书
人卫医学网	www.ipmph.com	医学考试辅导, 医学数 据库服务, 医学教育资 源, 大众健康资讯

版权所有, 侵权必究!

医学文献信息检索

第 3 版

主 编: 罗爱静 于双成

出版发行: 人民卫生出版社(中继线 010-59780011)

地 址: 北京市朝阳区潘家园南里 19 号

邮 编: 100021

E - mail: pmph@pmph.com

购书热线: 010-59787592 010-59787584 010-65264830

印 刷: 北京汇林印务有限公司

经 销: 新华书店

开 本: 850×1168 1/16 印张: 20

字 数: 550 千字

版 次: 2005 年 8 月第 1 版 2015 年 5 月第 3 版

2015 年 5 月第 3 版第 1 次印刷(总第 10 次印刷)

标准书号: ISBN 978-7-117-20571-9/R · 20572

定 价: 52.00 元

打击盗版举报电话: 010-59787491 E-mail: WQ@pmph.com

(凡属印装质量问题请与本社市场营销中心联系退换)

修订说明

为了贯彻教育部教高函[2004-9号]文,在教育部、原卫生部的领导和支持下,在吴阶平、裘法祖、吴孟超、陈灏珠、刘德培等院士和知名专家的亲切关怀下,全国高等医药教材建设研究会以原有七年制教材为基础,组织编写了八年制临床医学规划教材。从第一轮的出版到第三轮的付梓,该套教材已经走过了十余个春秋。

在前两轮的编写过程中,数千名专家的笔耕不辍,使得这套教材成为了国内医药教材建设的一面旗帜,并得到了行业主管部门的认可(参与申报的教材全部被评选为“十二五”国家级规划教材),读者和社会的推崇(被视为实践的权威指南、司法的有效依据)。为了进一步适应我国卫生计生体制改革和医学教育改革全方位深入推进,以及医学科学不断发展的需要,全国高等医药教材建设研究会在深入调研、广泛论证的基础上,于2014年全面启动了第三轮的修订改版工作。

本次修订始终不渝地坚持了“精品战略,质量第一”的编写宗旨。以继承与发展为指导思想;对于主干教材,从精英教育的特点、医学模式的转变、信息社会的发展、国内外教材的对比等角度出发,在注重“三基”、“五性”的基础上,在内容、形式、装帧设计等方面力求“更新、更深、更精”,即在前一版的基础上进一步“优化”。同时,围绕主干教材加强了“立体化”建设,即在主干教材的基础上,配套编写了“学习指导及习题集”、“实验指导/实习指导”,以及数字化、富媒体的在线增值服务(如多媒体课件、在线课程)。另外,经专家提议,教材编写委员会讨论通过,本次修订新增了《皮肤性病学》。

本次修订一如既往地得到了广大医药院校的大力支持,国内所有开办临床医学专业八年制及七年制(“5+3”一体化)的院校都推荐出了本单位具有丰富临床、教学、科研和写作经验的优秀专家。最终参与修订的编写队伍很好地体现了权威性、代表性和广泛性。

修订后的第三轮教材仍以全国高等学校临床医学专业八年制及七年制(“5+3”一体化)师生为主要目标读者,并可作为研究生、住院医师等相关人员的参考用书。

全套教材共38种,将于2015年7月前全部出版。

全国高等学校八年制临床医学专业国家卫生和计划生育委员会 规划教材编写委员会

名誉顾问

韩启德 桑国卫 陈 竺 吴孟超 陈灏珠

顾 问(按姓氏笔画排序)

马建辉 王 辰 冯友梅 冯晓源 吕兆丰 闫剑群 李 虹
李立明 李兰娟 杨宝峰 步 宏 汪建平 张 运 张灼华
陈国强 赵 群 赵玉沛 郝希山 柯 杨 桂永浩 曹雪涛
詹启敏 赫 捷 魏于全

主任委员

刘德培

委 员(按姓氏笔画排序)

丁文龙 于双成 万学红 马 丁 马 辛 丰有吉 王 杉
王兰兰 王宁利 王吉耀 王宇明 王怀经 王明旭 王建安
王建枝 王庭槐 王海杰 王家良 王鸿利 尹 梅 孔维佳
左 伋 冯作化 刘艳平 江开达 安 锐 许能锋 孙志伟
孙贵范 李 和 李 霞 李甘地 李明远 李桂源 李凌江
李继承 杨 恬 杨世杰 吴 江 吴忠道 何 维 应大君
沈 铿 张永学 张丽霞 张建中 张绍祥 张雅芳 陆 林
陈 红 陈 杰 陈孝平 陈建国 欧阳钦 尚 红 罗爱静
金征宇 周 桥 周 梁 赵旭东 药立波 柏树令 姜乾金
洪秀华 姚 泰 秦 川 贾文祥 贾弘禔 贾建平 钱睿哲
徐志凯 徐勇勇 凌文华 高兴华 高英茂 诸欣平 黄 钢
龚启勇 康德英 葛 坚 雷健波 詹希美 詹思延 廖二元
颜 虹 薛辛东 魏 泓

教材目录

	学科名称	主审	主编	副主编
1	细胞生物学(第3版)	杨 恬	左 伋 刘艳平	刘 佳 周天华 陈誉华
2	系统解剖学(第3版)	柏树令 应大君	丁文龙 王海杰	崔慧先 孙晋浩 黄文华 欧阳宏伟
3	局部解剖学(第3版)	王怀经	张绍祥 张雅芳	刘树伟 刘仁刚 徐 飞
4	组织学与胚胎学(第3版)	高英茂	李 和 李继承	曾园山 周作民 肖 岚
5	生物化学与分子生物学(第3版)	贾弘提	冯作化 药立波	方定志 焦炳华 周春燕
6	生理学(第3版)	姚 泰	王庭槐	闫剑群 郑 煜 祁金顺
7	医学微生物学(第3版)	贾文祥	李明远 徐志凯	江丽芳 黄 敏 彭宜红 郭德银
8	人体寄生虫学(第3版)	詹希美	吴忠道 诸欣平	刘佩梅 苏 川 曾庆仁
9	医学遗传学(第3版)		陈 竺	傅松滨 张灼华 顾鸣敏
10	医学免疫学(第3版)		曹雪涛 何 维	熊思东 张利宁 吴玉章
11	病理学(第3版)	李甘地	陈 杰 周 桥	来茂德 卞修武 王国平
12	病理生理学(第3版)	李桂源	王建枝 钱睿哲	贾玉杰 王学江 高钰琪
13	药理学(第3版)	杨世杰	杨宝峰 陈建国	颜光美 臧伟进 魏敏杰 孙国平
14	临床诊断学(第3版)	欧阳钦	万学红 陈 红	吴汉妮 刘成玉 胡申江
15	实验诊断学(第3版)	王鸿利 张丽霞 洪秀华	尚 红 王兰兰	尹一兵 胡丽华 王 前 王建中
16	医学影像学(第3版)	刘玉清	金征宇 龚启勇	冯晓源 胡道予 申宝忠
17	内科学(第3版)	王吉耀 廖二元	王 辰 王建安	黄从新 徐永健 钱家鸣 余学清
18	外科学(第3版)		赵玉沛 陈孝平	杨连粤 秦新裕 张英泽 李 虹
19	妇产科学(第3版)	丰有吉	沈 铿 马 丁	狄 文 孔北华 李 力 赵 霞

	学科名称	主审	主编	副主编
20	儿科学(第3版)		桂永浩 薛辛东	杜立中 母得志 罗小平 姜玉武
21	感染病学(第3版)		李兰娟 王宇明	宁 琴 李 刚 张文宏
22	神经病学(第3版)	饶明俐	吴 江 贾建平	崔丽英 陈生弟 张杰文 罗本燕
23	精神病学(第3版)	江开达	李凌江 陆 林	王高华 许 毅 刘金同 李 涛
24	眼科学(第3版)		葛 坚 王宁利	黎晓新 姚 克 孙兴怀
25	耳鼻咽喉头颈外科学(第3版)		孔维佳 周 梁	王斌全 唐安洲 张 罗
26	核医学(第3版)	张永学	安 锐 黄 钢	匡安仁 李亚明 王荣福
27	预防医学(第3版)	孙贵范	凌文华 孙志伟	姚 华 吴小南 陈 杰
28	医学心理学(第3版)	姜乾金	马 辛 赵旭东	张 宁 洪 炜
29	医学统计学(第3版)		颜 虹 徐勇勇	赵耐青 杨士保 王 彤
30	循证医学(第3版)	王家良	康德英 许能锋	陈世耀 时景璞 李晓枫
31	医学文献信息检索(第3版)		罗爱静 于双成	马 路 王虹菲 周晓政
32	临床流行病学(第2版)	李立明	詹思延	谭红专 孙业桓
33	肿瘤学(第2版)	郝希山	魏于全 赫 捷	周云峰 张清媛
34	生物信息学(第2版)		李 霞 雷健波	李亦学 李劲松
35	实验动物学(第2版)		秦 川 魏 泓	谭 毅 张连峰 顾为望
36	医学科学研究导论(第2版)		詹启敏 王 杉	刘 强 李宗芳 钟晓妮
37	医学伦理学(第2版)	郭照江 任家顺	王明旭 尹 梅	严金海 王卫东 边 林
38	皮肤性病学	陈洪铎 廖万清	张建中 高兴华	郑 敏 郑 捷 高天文

第三版序言

经过再次打磨,备受关爱期待,八年制临床医学教材第三版面世了。怀纳前两版之精华而愈加求精,汇聚众学者之智慧而更显系统。正如医学精英人才之学识与气质,在继承中发展,新生方可更加传神;切时代之脉搏,创新始能永领潮头。

经过十年考验,本套教材的前两版在广大读者中有口皆碑。这套教材将医学科学向纵深发展且多学科交叉渗透融于一体,同时切合了环境-社会-心理-工程-生物这个新的医学模式,体现了严谨性与系统性,诠释了以人为本、协调发展的思想。

医学科学道路的复杂与简约,众多科学家的心血与精神,在这里汇集、凝结并升华。众多医学生汲取养分而成长,万千家庭从中受益而促进健康。第三版教材以更加丰富的内涵、更加旺盛的生命力,成就卓越医学人才对医学誓言的践行。

坚持符合医学精英教育的需求,“精英出精品,精品育精英”仍是第三版教材在修订之初就一直恪守的理念。主编、副主编与编委们均是各个领域内的权威知名专家学者,不仅著作立身,更是德高为范。在教材的编写过程中,他们将从医执教中积累的宝贵经验和医学精英的特质潜移默化地融入到教材中。同时,人民卫生出版社完善的教材策划机制和丰富的编辑队伍保障了教材“三高”(高标准、高起点、高要求)、“三严”(严肃的态度、严谨的要求、严密的方法)、“三基”(基础理论、基本知识、基本技能)、“五性”(思想性、科学性、先进性、启发性、适用性)的修订原则。

坚持以人为本、继承发展的精神,强调内容的精简、创新意识,为第三版教材的一大特色。“简洁、精练”是广大读者对教科书反馈的共同期望。本次修订过程中编者努力做到:确定系统结构,落实详略有方;详述学科三基,概述相关要点;精选创新成果,简述发现过程;逻辑环环紧扣,语句精简凝练。关于如何在医学生阶段培养创新素质,本教材力争达到:介绍重要意义的医学成果,适当阐述创新发现过程,激发学生创新意识、创新思维,引导学生批判地看待事物、辩证地对待知识、创造性地预见未来,踏实地践行创新。

坚持学科内涵的延伸与发展,兼顾学科的交叉与融合,并构建立体化配套、数字化的格局,为第三版教材的一大亮点。此次修订在第二版的基础上新增了《皮肤性病学》。本套教材通过编写委员会的顶层设计、主编负责制下的文责自负、相关学科的协调与磋商、同一学科内部的专家互审等机制和措施,努力做到其内容上“更新、更深、更精”,并与国际紧密接轨,以实现培养高层次的具有综合素质和发展潜能人才的目标。大部分教材配套有“学习指导及习题集”、“实验指导/实习指导”以及“在线增值服务(多媒体课件与在线课程等)”,以满足广大医学院校师生对教学资源多样化、数字化的需求。

本版教材也特别注意与五年制教材、研究生教材、住院医师规范化培训教材的区别与联系。①五年制教

材的培养目标:理论基础扎实、专业技能熟练、掌握现代医学科学理论和技术、临床思维良好的通用型高级医学人才。②八年制教材的培养目标:科学基础宽厚、专业技能扎实、创新能力强、发展潜力大的临床医学高层次专门人才。③研究生教材的培养目标:具有创新能力的科研型和临床型研究生。其突出特点:授之以渔、评述结合、启示创新,回顾历史、剖析现状、展望未来。④住院医师规范化培训教材的培养目标:具有胜任力的合格医生。其突出特点:结合理论,注重实践,掌握临床诊疗常规,注重预防。

以吴孟超、陈灏珠为代表的老一辈医学教育家和科学家们对本版教材寄予了殷切的期望,教育部、国家卫生和计划生育委员会、国家新闻出版广电总局等领导关怀备至,使修订出版工作得以顺利进行。在这里,衷心感谢所有关心这套教材的人们!正是你们的关爱,广大师生手中才会捧上这样一套融贯中西、汇纳百家的精品之作。

八学制医学教材的第一版是我国医学教育史上的重要创举,相信第三版仍将担负我国医学教育改革的使命和重任,为我国医疗卫生改革,提高全民族的健康水平,作出应有的贡献。诚然,修订过程中,虽力求完美,仍难尽人意,尤其值得强调的是,医学科学发展突飞猛进,人们健康需求与日俱增,教学模式更新层出不穷,给医学教育和教材撰写提出新的更高的要求。深信全国广大医药院校师生在使用过程中能够审视理解,深入剖析,多提宝贵意见,反馈使用信息,以便这套教材能够与时俱进,不断获得新生。

愿读者由此书山拾级,会当智海扬帆!

是为序。

中国工程院院士

中国医科科学院原院长

北京协和医学院原院长

劉德培

二〇一五年四月

主编简介

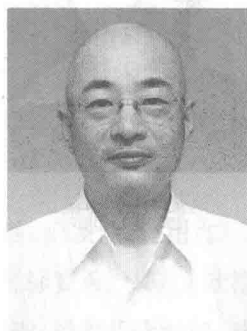
罗爱静,教授,博士生导师,享受国务院政府特殊津贴专家,湖南省“121工程”人才,中南大学湘雅三医院党委书记,中南大学第一届、第二届党委委员。兼任中国卫生信息学会卫生信息学教育专业委员会副主任委员,湖南省医疗机构信息管理质量控制中心主任,湖南省高等学校医学信息研究重点实验室主任,全国医学文检教研会副理事长,中华医学会数字医学分会委员,湖南省科技情报学会副理事长。国家精品课程《文献信息检索》负责人,主持了国家社科基金2项,国家科技惠民计划项目、教育部人文社会科学研究项目、教育部博士点基金项目、卫生部项目等30余项;发表论文90余篇;主编8年制临床医学专业规划教材等国家级规划教材多部;研究成果获国家五部委科技情报优秀成果奖,湖南省科技进步三等奖,湖南省哲学社会科学优秀成果三等奖,湖南省教学成果二等奖,中国图书馆学会、湖南省科技情报学会一等奖等奖项20余项;湖南省科技信息工作突出贡献奖获得者。



罗爱静

于双成,教授,吉林大学教学示范教师、吉林省优秀教学团队负责人、吉林省省级精品课负责人。任《医学与哲学》和《临床肝胆病杂志》编委、全国医学文献检索教学研究会副理事长、全国高等学校八年制临床医学专业卫生部规划教材编委会委员、教育部人文社会科学课题(教育学)评审专家、教育部医学人文素质教学指导委员会委员。

在医学方法学、医学信息检索、医学教育等领域发表学术论文120余篇,主编(参编)教科书、专著20余部。



于双成

副主编简介



马 路

马路,教授,硕士生导师。现任首都医科大学图书馆馆长、档案馆馆长,首都医科大学卫生管理与教育学院卫生与医学信息管理学系主任。教育部高等学校图书情报工作指导委员会委员、中国图书馆学会学术研究委员会委员、中国中西医结合学会信息专业委员会常务委员、中华预防医学会预防医学信息专业委员会委员、中华医学会医学信息学分会委员会委员。

从事教学工作二十余年,主要研究方向为文献资源建设、图书馆管理。主编、副主编教材9部,发表论文数十篇。



王虹菲

王虹菲,研究馆员,大连医科大学图书馆馆长,辽宁省第一、二届医学信息学会副主任委员、大连图书馆学会副理事长、辽宁图书馆学会理事、大连医学会医学信息专科分会前任主委。

从事教学工作25年,编写相关教材10余部、发表专业论文20余篇。主要研究方向为图书馆管理学、数据库建设、医学文献检索等。主持和参与各类课题5项,获得辽宁省科技奖励三等奖。



周晓政

周晓政,研究馆员,江苏省高校图工委情报咨询专业委员会委员。从事医学图书馆知识集整合整理、传播工作30余年,1984年起从事医学文献信息检索与利用课程教学至今。2012年长三角优秀科技情报工作者。主要作品有《疾病改变历史》(2004年合译)《太平洋战场》(1996年合著)《医学信息检索教程》(2002年合著)《医药信息检索与利用》(2006年主编)《医药信息检索与利用》(2012年主编)《医学影像信息学》(2014年合著)《医药学文献计算机检索:光盘、Internet联机信息资源和联机检索方法及技巧》(1999年副主编)《医学文献信息检索(2版)》(2010年参编)等。

2004年,我国开始在部分高校试办八年制医学教育,这种学制是一种新的临床医学专业培养模式,也是目前我国最高层次的医学教育,其目的是与国际接轨、培养一批具有医学博士学位的高层次、高素质的、具有国际竞争力的医学精英人才。八年制医学教育注重综合能力的培养:包括自学能力,动手能力,思维能力,创新能力,理论联系实际能力,独立提出、分析、解决问题的能力,协调沟通能力等。在当今信息化时代,这种综合能力的形成需要有较高的信息素养作为坚实的基础。对八年制医学生而言,医学信息素养的形成与发展,不仅关系到学业,而且关系到毕业后的学术生涯。

医学文献信息检索课程是一门培养学生的信息意识,掌握从浩如烟海的文献中获取知识和信息的技能,提高学生自学能力和独立研究问题、解决问题能力,提高学生的信息素养的方法课。因而,在培养高素质、厚基础、宽口径、创新能力强和有独立研究能力的高层次复合型人才的过程中,信息检索课程具有举足轻重的作用。

本教材第2版发行5年以来,在全国开办长学制的高等医药院校得到了广泛使用,也得到了广大师生的一致好评。但是随着医学的发展和信息环境的变化,亟待丰富与更新,将一些新概念、新方法、新技术和新观念融入其中,以适应时代发展的需要、适应医学教育的需要、适应培养医学精英人才的需要。

本版教材作为我国高等学校临床医学专业八年制国家卫生和计划生育委员会规划教材的第3版,继承和发扬了第2版的许多优点,同时又补充和增添了一些新的内容,立足培养学生信息素养,开拓知识共享空间,为新时代的医学精英教育作出贡献。在本版教材的编写过程中,我们努力做到富于创新、敢于挑战,力求突破。

1. 坚持“三基”(基础理论、基本知识、基本技能)、“五性”(思想性、科学性、先进性、启发性、适用性) 全面系统介绍医学文献信息检索的基本原理、基本理论、技术方法。基于精英教育的目标要求,体现文献检索课方法学课程的特点,本着理论与实践并重,深度与广度兼顾,检索与利用互动的原则,沿着检索知识、检索技能、信息思维这一轴线选择和组织内容。

2. 重视“信息素养”教育 在全球信息化的背景下,“信息素养(information literacy)”是一种适应社会的基本能力,也是全球医学教育的基本要求,医学文献信息检索课程将以提高医学生的“信息素养”为重要目标,培养高层次的具有综合素质和发展潜能的人才。本版教材将增加当前国内外信息素养的教育实践情况,从中汲取开展信息素养教育的宝贵经验,探索中国式信息素养教育之路。

3. 内容更新 教材内容不仅“更新、更深、更精”,适时反映信息检索理论与技术的最新进展,力求全面、及时、准确反映迅速增长的医学科技信息资源的新变化,以及最新的生物医学信息检索技术、理论与方法;而且紧随信息技术日新月异的步伐和生物医学科学的迅猛发展,增加了一些新的独到内容,如移动图书馆、知识管理工具、国外信息素养教育平台、转化医学信息、知识型数据库、新的医学视频资源库以及移动网络资源搜索等。

4. 注重编排体例和知识安排的创新 教材编排上循序渐进,理论与实际相结合,增加并统一分析案例,采用以问题为基础的教学方法,更加注重启发性和实用性。

5. 加强多媒体课件与在线课堂视频 本版教材增加多媒体与在线课堂视频的制作,充分发挥多媒体的优势,较多地采用图片、动画等多媒体素材,从而帮助学生更快捷直观地掌握中外文医学文献资源等内容的检索方法,增加趣味性。

6. 完善配套教材 随书配套学习指导与习题集,增加案例教学内容,拓展课后阅读,开拓学习思维能力,强化学习效果,增强学生课后复习的针对性和启发性。

本教材的编委团队在原有基础上吸入了新鲜血液,更为权威,更具代表性,云集了北京大学、北京协和医科大学、吉林大学、复旦大学、上海交通大学、中山大学、华中科技大学、中南大学、首都医科大学、中国医科大学、重庆医科大学、大连医科大学、哈尔滨医科大学、南京医科大学、江苏大学、山西医科大学等16所高校的专家教授,遵循创新、融合的原则,聚集了各专家的智慧、经验与最新成果,可谓“精英出精品,精品育精英”。在此,对各位编者所付出的辛勤劳动表示感谢。本书也得到了国家卫生计生委教材办公室、人民卫生出版社、大连医科大学图书馆、中南大学湘雅三医院、湘雅医学院医药信息系和各位编者所在单位的大力支持,在此表示诚挚的谢意。同时,感谢被引用的各种参考文献的作者,是他们的研究成果为本教材增光添彩。奠定了本教材的编写基础。

随着网络信息资源的变化更新,文献信息管理科学及计算机技术的快速发展,许多学科和实践课题还有待进一步探讨。囿于我们的水平,书中难免存在缺点和不足,恳请读者指正和赐教。

罗爱静 于双成

2015年5月

目 录

绪论	1
第一节 课程的作用、目标与任务	1
一、医学人才的知识能力要求	1
二、医学文献信息检索课程的作用	2
三、医学文献信息检索课程的性质、目标与任务	4
第二节 信息素养及医学信息素养	5
一、信息素养的概念	5
二、信息素养教育及评价标准	6
三、医学信息素养内涵及发展	7
第一章 文献信息检索基础	12
第一节 文献信息检索概述	12
一、文献信息概述	12
二、文献信息检索概述	18
第二节 文献信息检索系统	20
一、文献信息检索系统的构成	21
二、文献信息数据库的结构	21
三、文献信息检索系统的分类	23
四、文献信息检索系统的功能	24
第三节 文献信息检索技术	24
一、文本信息检索技术	25
二、多媒体信息检索技术	27
三、可视化信息检索技术	28
第四节 文献信息检索途径与策略	29
一、文献信息检索途径	29
二、文献信息检索策略	30
第二章 文献信息检索语言	34
第一节 文献信息检索语言概述	34
一、检索语言的概念	34
二、检索语言的构成	34
三、检索语言的作用	35
四、检索语言的类别	35

第二节 文献分类法	37
一、文献分类法概述	37
二、常用文献分类法	38
第三节 国际疾病分类法	41
一、国际疾病分类法概述	41
二、学习国际疾病分类法的意义	42
三、国际疾病分类法的主体结构	42
第四节 国际系统医学学术语集	44
一、发展沿革及应用	44
二、概念体系	45
第五节 医学主题词表	48
一、概述	48
二、概念体系	48
三、MeSH Browser	50
四、一体化医学语言系统	51
第三章 图书馆信息资源利用	54
第一节 馆藏资源的获取	54
一、馆藏目录查询	54
二、资源发现系统	55
三、移动图书馆	55
第二节 图书馆信息服务	57
一、文献借阅	57
二、参考咨询与读者培训	57
三、科技查新	57
四、馆际互借与文献传递	57
五、学科服务	58
六、信息共享空间	58
第三节 信息资源共享	58
一、CALIS 资源与服务	58
二、NSTL 资源与服务	59
第四节 学术期刊与参考工具书	59
一、学术期刊	59
二、参考工具书	60
第五节 医学电子图书	62
一、电子图书概述	62
二、电子图书举例	63
第四章 中文数据库检索	66
第一节 中国生物医学文献服务系统	66
一、概述	66

二、检索方法和技巧	67
三、检索结果的处理	71
四、个性化服务	71
第二节 中国医院知识总库(CKHD)	71
一、概述	71
二、检索方法和技巧	72
三、检索结果的处理	75
第三节 万方医学网	75
一、概述	75
二、检索方法和技巧	76
三、检索结果的处理	77
第四节 维普期刊资源整合服务平台	78
一、概述	78
二、检索方法和技巧	78
三、检索结果的处理	79
第五章 外文文摘数据库检索	81
第一节 PubMed	81
一、概述	81
二、检索方法与技巧	83
三、检索结果的处理	88
四、个性化功能与服务	90
第二节 Embase	92
一、概述	92
二、检索方法与技巧	93
三、检索结果的处理	98
四、个性化功能与服务	99
第三节 BIOSIS Previews	99
一、概述	99
二、检索方法与技巧	100
三、检索结果的处理	102
四、个性化功能与服务	103
第四节 SciFinder Scholar	103
一、概述	103
二、检索方法与技巧	103
三、检索结果的处理	106
四、个性化功能与服务	108
第六章 外文全文数据库检索	110
第一节 ClinicalKey	110
一、概述	110

二、检索方法与技巧	111
三、检索结果的处理	114
四、个性化功能与服务	115
第二节 Journals@ovid Full Text	116
一、概述	116
二、检索方法与技巧	116
三、检索结果的处理	118
四、个性化功能与服务	119
第三节 Wiley Online Library	120
一、概述	120
二、检索方法与技巧	120
三、检索结果的处理	122
四、个性化功能与服务	123
第四节 SpringerLink	124
一、概述	124
二、检索方法与技巧	125
三、检索结果的处理	126
四、个性化功能与服务	127
第五节 其他外文全文数据库	127
一、ScienceDirect	127
二、CINAHL PLUS with Full Text	129
三、ProQuest	132
四、Karger	136
第七章 引文检索	140
第一节 引文检索概述	140
一、基本概念	140
二、引文检索的作用	140
第二节 Web of Science 核心合集	141
一、概述	141
二、检索规则	142
三、检索方法和技巧	142
四、检索结果处理和分析	146
五、个性化功能和服务	148
第三节 中文引文检索	149
一、中国科学引文数据库	149
二、中国引文数据库	151
三、其他中文数据库的引文检索	153
第四节 引文分析工具	156
一、期刊引用报告	156
二、基本科学指标	159

第八章 专类信息检索	161
第一节 学位论文检索	161
一、概述	161
二、国内学位论文检索	161
三、国外学位论文检索	163
第二节 会议文献检索	164
一、概述	164
二、会议预告信息检索	165
三、国内会议论文检索	166
四、国外会议论文检索	167
第三节 专利文献检索	168
一、概述	168
二、国内专利检索	170
三、国外专利检索	176
第四节 循证医学资源检索	181
一、概述	181
二、循证医学资源数据库检索	184
第五节 药学信息检索	191
一、综合性药学文献数据库	191
二、与临床处方相关的药物信息	194
三、与药物开发相关的信息	196
四、机构信息	199
五、综合性药学专业网站	200
第六节 疾病知识库检索	202
一、中文疾病知识库	203
二、国外疾病知识库	205
第七节 医学视频数据库检索	208
一、万方医学视频数据库	208
二、Primal Pictures 人体解剖 3D 模型数据库	209
三、美国视频实验期刊(JoVE)	209
四、爱思唯尔 ClinicalKey 医学视频库	210
第九章 生物信息数据库检索	212
第一节 生物信息数据库概述	212
一、生物信息数据库的类型	212
二、生物信息数据收集与存储	213
三、生物信息数据库的查找	215
第二节 核酸序列数据库	215
一、GenBank	215
二、ENA	217
三、DDBJ	218

第三节 蛋白质数据库	220
一、蛋白质序列数据库	220
二、蛋白质结构数据库	222
三、蛋白质功能数据库	226
第四节 基因组数据库	227
一、Entrez Genomes	227
二、Ensembl	229
三、UCSC Genome Browser	231
第五节 疾病基因数据库	233
一、OMIM	233
二、GeneCards	235
三、GEO	235
四、HGMD	236
五、单核苷酸多态性数据库	236
第十章 网络信息资源检索	239
第一节 网络信息资源概述	239
一、网络信息资源的定义	239
二、网络信息资源的特点	239
第二节 搜索引擎	240
一、概述	240
二、Google Scholar 和百度学术搜索	241
三、医学搜索引擎	241
四、搜索引擎的使用技巧	243
第三节 开放存取资源	244
一、概述	244
二、国外主要开放存取资源	245
三、国内主要开放存取资源	248
四、慕课资源	249
第十一章 文献信息分析与利用	252
第一节 个人文献信息的搜集与整理	252
一、个人文献信息的搜集	252
二、个人文献信息的整理和积累	253
三、文献信息的评价	255
第二节 文献信息分析	259
一、文献信息分析概述	259
二、文献信息分析的步骤	259
三、文献信息分析的方法	263
第三节 文献信息的利用	274
一、文献信息利用的层次与基本原则	274

二、科学研究中的文献信息利用	275
三、管理活动中文献信息利用	276
附录.....	279
一、MeSH 词表树状结构体系的大类和二级类目 (2015)	279
二、MeSH 词表限定词使用一览表 (2015)	282
参考文献及推荐网站.....	289
中英文名词对照索引	291
致谢.....	295

绪 论

医学科学的发展和医疗服务的深化,均以有效获取与利用现有知识信息为前提。因此,统领和驾驭医学文献信息的信息素养被视为高素质医学人才形成和发展的重要基础。医学文献信息检索,作为一门科学方法课程,着力于培养医学生的信息意识,获取文献信息方法和技能,以及蕴含于信息行为活动中的信息道德修养。

第一节 课程的作用、目标与任务

一、医学人才的知识能力要求

世界卫生组织阿拉木图宣言和世界医学教育爱丁堡宣言均指出:医学教育的最高标准是满足社会的需要。1999年6月9日,经纽约中华医学基金会(China Medical Board of New York, CMB)理事会批准资助,成立了国际医学教育专门委员会(Institute for International Medical Education, IIME)。IIME通过收集和 research 世界各国医学教育的要求和标准,制定出全球医学教育最基本标准,即“全球医学教育最低基本要求(global minimum essential requirement medical education)”,它主要包括7个宏观领域:职业价值、态度、行为和伦理;医学科学基础知识;沟通技能;临床技能;群体健康和卫生系统;信息管理;批判性思维和研究。

医学教育具有精英教育的性质,在国际上医学教育多以长学制教育为主。医学生的来源主要有两种:一种是以美国为代表的北美洲国家,这些国家的医学生均来自自己完成3~4年大学本科教育的预科生或毕业生;一种是以法国、英国、德国、俄国和日本为代表的国家,这些国家的医学生一般均从高中毕业生中选拔,实行一贯制教育。不同层次医学院校的培养目标在总的人才培养目标下也不同,如哈佛大学培养目标:创造和培育优秀的多元化社会人才,致力于减轻人们因疾病所造成的痛苦;杜克大学培养目标:培养未来的临床和科研领导型人才。

随着我国医疗卫生事业的不断发展,社会对医学人才的要求越来越高,需要更注重交叉、前沿、新兴学科知识掌握与应用的高层次医学人才。近年来,我国实行了一系列高等医学教育改革措施,在如何进一步提高医学人才的培养规格和培养质量方面进行了有益的探索。2004年我国恢复试办八年制医学教育,这种学制代表了当前医学教育的最高层次,学生毕业时可获得医学博士(medical doctor, M.D.)学位。经过这种模式培养出来的毕业生应是高素质、厚基础、宽口径、创新能力强和有独立研究能力的高层次复合型人才。

国内一流高等医学院校借鉴发达国家高等医学教育培养模式培养八年制医学博士已成为医学教育发展的主导方向。根据教育部提出的“八年一贯、整体优化、强化基础、注重临床、培养能力、提高素质”培养原则,各院校都制定了八年制临床医学的培养目标,提出了高素质医学人才的知识能力结构要求。

中国协和医科大学具有悠久的八年制教学经验,坚持医学精英教育,注重综合能力的培养,实行“高进”“优教”“严出”;强调“三高”“三基”“三严”,博学众长,开放办学,培养高层次的具有综合素质和发展潜能的人才。北京大学、复旦大学、中南大学、中山大学、华中科技大学等十余所院校开设了八年制医学教育,坚持与国际教育接轨,分别制定了八年制医学人才的培养目标。

培养目标强调了以下几个方面:①有较宽厚的人文和自然科学素养;②有较高的医学理论水平;③有较高的临床医学实践能力和科研素质;④有较强的自我发展能力和发展潜力;⑤有很强的社会适应性,有较好的沟通技能和与他人(包括国际)交流、协作能力;⑥有正确的职业价值观和较高的综合素质。也就是说,我国高素质医学人才的知识能力结构要求要包括以上六方面。

根据高素质医学人才的知识能力结构要求,各高校也探索了不同的培养模式。中南大学湘雅医学院推行医学前教育模式,将所有的课程整合为“思想政治与人文社会科学”“自然科学基础”“生命科学基础”“医学专业与科研训练”和“身心素质”五大模块。中山大学医学院利用优势教学资源,与全球医学教育最基本要求相适应,构建了“243”型课程体系:即核心课程+选修超市的2重格局;课程组建体现4个主题:健康与疾病状态下的身体机能运作、医生与患者(患者家属)的关系、医学与社会(生理/心理/环境)的关系、专业与专业技能发展的关系;教学内容和教学方法加强3个渗透:健康人格培养与科学知识教育相渗透、医学人文与方法相渗透、基础与临床相渗透。中国医科大学在长期实践中总结出以培养创新精神与实践能力为核心的医学人才培养模式:一厚:宽厚的理工基础教育;二早:早期接触临床、早期科研训练;三结合:人文教育与医学科学紧密结合、基础学科与临床医学紧密结合、临床医学与预防医学紧密结合;四提高:在学生学习的七年中逐渐提高职业道德观念、提高外语能力、提高实验创新能力、提高临床实践能力。

二、医学文献信息检索课程的作用

高等医学人才的培养任重而道远,随着社会信息化的发展,医学文献信息检索课程在高等医学人才的综合素质培养中起着举足轻重的作用。

(一) 医学文献信息检索课程与信息素养

“全球医学教育最低基本要求”是培养医学人才的具体目标和要求,是我国高等医学院校教学评估的标准,也为我国医学教育改革的发展提供了方向。毋庸置疑,“全球医学教育最低基本要求”也是对医学人才信息素质教育的新要求。首先,在“沟通技能”这一宏观领域中提出为提高医疗决策的准确性和患者的满意度,医学毕业生必须能够专心倾听,以收集和综合与各种问题有关的信息,并且理解这些内容,有效地与同事、教师、社区、其他部门以及公共媒体之间进行沟通和交流,有效地进行口头和书面的沟通,能综合并向听众介绍适合他们需要的信息,与他们讨论关于解决个人和社会重要问题的可达到的和可接受的行动计划。其次,在“信息管理”中提出医学毕业生必须有能力从不同的数据库中检索、搜集、组织和分析有关卫生和生物医学的信息,从临床医学数据库中检索特定患者的信息,运用信息和通信技术帮助疾病诊断、治疗和预防,以及对健康的调查和监控,懂得信息技术的运用及其局限性以及保存医疗工作的记录,以便进行分析和改进。再者,在“批判性思维”中着重强调医学毕业生在职业活动中能表现出分析批判的精神、有根据的怀疑、创造精神和对事物进行研究的态度,懂得根据从不同信息源获得的信息在确定疾病的病因、治疗和预防中进行科学思维的重要性和局限性,应用个人判断来分析和评论问题,主动寻求信息,根据相关信息,运用科学思维去识别、阐明和解决患者的问题,理解在做出医疗决定时应考虑到问题的复杂性、不确定性,提出假设,收集并评价各种资料,从而解决问题。

21世纪全球高等教育的理念正由知识教育向素质教育转变,信息素质教育作为人的其他各项素质的基础,已成为新世纪人才所需的基本素质。世界各国都把信息素质教育作为教育发展的重要推动力,如美国推出教育技术创新计划,澳大利亚实施教育网计划,日本进行中小学联网试验,韩国开展“教育革命”等。美国国家科学顾问委员会在1999年2月提出的《新的千纪年信息技术前沿蓝皮书》指出:21世纪人类对于高性能的运算和通信的要求将日益增长,培养

每个公民使用以计算机为基础的网上获取、处理和交换信息的能力将是未来教育的主要任务。

20 世纪 80 年代,我国就已经开始重视信息教育,普及信息检索知识。1984 年,教育部下发“在全国高校开展文献检索与应用课”的文件,从而引起全国各高校开设文献检索课的热潮。2002 年 1 月,教育部高等学校图书情报工作指导委员会在黑龙江大学召开全国高校信息素养教育学术研讨会,会议呼吁各高校重视信息素养教育,提出要加快改革以文献课教学为主体的信息素养教育课程,加强教材建设。

信息素养(information literacy)是指判断何时需要信息,并有效地定位、获取、评价和利用信息并致力于终身学习的一系列能力的总和。培养和提高医学生的信息素养,必须使每位医学生都善于检索和利用信息与知识,实现信息和知识效用的最大化。因此,培养高级医学人才的信息素养,提高医学人才的信息意识、信息检索能力、信息分析能力、信息利用能力等,是我们开展医学文献信息检索课的重要任务。

(二) 医学文献信息检索课程与终身教育

终身教育(lifelong education)是 20 世纪 50 年代末 60 年代初形成于欧洲的一种国际教育思潮,也是至今依然对国际教育事业产生重大影响的一种新的教育观念和教育理论。它不仅从宏观上影响着世界各国教育战略的规划,而且从微观上涉及教育理论和实践的各个具体方面。1972 年,由联合国教科文组织召集,法国前总理埃德加·富尔主持的国际委员会上提出了题为《学会生存:教育世界的今天和明天》的报告。该报告正式确认了原教科文组织终身教育处处长保罗·朗格让于 20 世纪 60 年代中期提出的终身教育理论。联合国教科文组织还提出,教育已扩大到一个人的整个一生,认为只有全面的终身教育才能够培养完善的人,可以防止知识老化,不断更新知识,适应当代信息社会发展的需求。近年来,终身教育观念日益深入人心。我国教育部在《教育信息化十年发展规划(2011—2020 年)》中提出,继续教育信息化是建设终身学习体系的重要支撑,为实现我国继续教育信息化的重要目标,分三个层次推进继续信息化教育:①推进继续教育数字资源建设与共享;②加快信息化终身学习公共服务体系建设;③加强继续教育公共信息管理与服务平台建设,推进开放大学建设,面向全社会提供服务,为学习者提供方便、灵活、个性化的信息化学习环境,促进终身学习体系和学习型社会建设。

信息检索是终身教育的基础。学校培养学生的目标是学生的智能:包括自学能力、研究能力、思维能力、表达能力和组织管理能力。在知识经济时代,知识飞速发展,科技日新月异,知识的时效性、知识增长和扩展的无限性,决定了学无止境。医学是一个知识密集型的学科,更应该时时更新所学知识,以适应现实需求。

医学文献信息检索与终身教育息息相关。具备医学文献信息检索能力的医学生进入社会后,在工作中可以及时跟踪、掌握医学科技发展的最新动态,获取网上各种重要医学信息和最新医疗手段,拓展知识面,调整知识结构,创新诊疗工作。因而,医学文献信息检索素质是学生投身社会后延伸教育和终身教育的基础。

(三) 医学文献信息检索课程与创新型人才培养

所谓创新型人才,是指富于独创性,具有创新意识、创新精神、创新思维和创新能力,具有能够提出、分析和解决问题的能力,能够开创事业新局面,对社会物质文明和精神文明建设作出创造性贡献的人。习近平同志在 2014 年 10 月中国科学院第十七次院士大会、中国工程院第十二次院士大会上强调:“我国科技发展的方向就是创新、创新、再创新。要坚定不移走中国特色自主创新道路,坚持自主创新、重点跨越、支撑发展、引领未来的方针,加快创新型国家建设步伐。科技是国家强盛之基,创新是民族进步之魂。创新的事业呼唤创新的人才。实现中华民族伟大复兴,人才越多越好,本事越大越好。知识就是力量,人才就是未来。我国要在科技创新方面走在世界前列,必须在创新实践中发现人才、在创新活动中培育人才、在创新事业中凝聚人才,必须大力培养造就规模宏大、结构合理、素质优良的创新型科技人才。”高等院校肩负

着培养和造就具有创造性思维人才的历史使命。创造性思维是指人在实践学习中,根据自己的目标展示出来的一种主动、独创、富有新颖特点的思维方式,它是在原有经验和学得知识的基础上进行合理性和突破性的创造组合,形成新的概念或新的成果。创新型医学人才应具有医学认知能力和医学创新能力。医学认知能力是指以良好的医学知识结构为基础、以医学的科学精神和人文情怀为准绳、以正确方式方法的运用为前提、以发现和解析问题为特征的,揭示生命本质与疾病规律的认知能力。医学创新能力是指创新医学知识、发展医学科学的认知能力,是基于医学认知能力基础之上的进一步升华。医学认知能力和医学创新能力的形成都离不开医学文献信息检索等医学方法学课程的支撑。

在我国,传统教学以传授已有知识为目的,多采用“填鸭式”“注入式”教学方法,强调对知识的记忆、模仿和重复练习,极易束缚和压抑学生的创新精神。为适应创新型社会的发展和科技进步的要求,教育部制定了《国家中长期教育改革与发展规划纲要(2010—2020年)》,提出“到2020年基本实现教育现代化,基本形成学习型社会,进入人力资源强国”的战略目标;该纲要还强调学思结合,倡导启发式、探究式、讨论式、参与式教学,帮助学生学会学习。目前,常见的启发式教学法有:PBL教学、案例式教学、网络开放式教学等。这些方法有利于培养学生的创新意识,与医学文献信息检索能力关联密切。

美国医学教育提倡CBL(以社区为基础的学习方法)和PBL(以问题为中心的学习方法)等教育方法,其主要特征为:①以突出问题为中心,以病例为基础的讨论式小组学习;②明确教学目标的自我教育和主动学习;③加强医患关系与社区卫生知识、技能的学习锻炼。在这些教育方法中,老师激发与调动学生最大限度的学习积极性,提高学生的应用知识、解决问题、获取与掌握知识的能力,促进人际交流与介入社会的技能。启发式教学改变传统的以教师为中心的教学方法,重在引导,通过师生、学生之间的互动,平等对话和研讨,提高学生学习的兴趣,增强学生分析问题与解决问题能力、交流能力、合作能力、语言表达能力和理论联系实际的能力。启发式教学方法成功推行,需要学生课前进行大量的查找资料、收集资料的准备工作。医学文献信息检索课程旨在培养学生信息意识、信息能力、信息道德,以及学生的检索、筛选、分析、评价、管理和综合利用文献信息的能力,创新能力和终身学习能力。因而,通过医学文献信息检索课程的学习,有利于启发式教学的顺利开展;同时,启发式教学促进了学生对医学文献信息检索课程学习内容的实际应用,增强了学生对医学文献信息检索课程的重视度,提高了学生学习的主动性和积极性。

启发式教学方法和医学文献信息检索课程相互渗透,扩大学生视野,培养学生创新性思维,从而推动我国创新型医学人才的培养工作。

三、医学文献信息检索课程的性质、目标与任务

原国家教委曾分别于1984年、1985年、1992年3次以文件的形式督促各高校开设文献检索与利用课,并指出该课程的目的是“使大学生和研究生增强情报意识,初步掌握利用文献与情报技能”,该课程“对人们不断吸收新知识,改善知识结构、提高自学能力和研究能力,发挥创造才能都具有重要的意义”。在2002年新修订的《普通高等学校图书馆规程》中也规定图书馆“通过开设文献信息检索与利用课程以及其他手段,进行信息素质教育”。因而,医学文献信息检索课程是一门非常重要的课程。

(一) 医学文献信息检索课的性质

医学文献信息检索隶属于医学方法学范畴,而医学方法学是指源自多学科的方法学知识,分别与医学专业知识相融合而形成的,有助于医学知识获取、利用与产出的工具性学科。医学文献信息检索,正是一门融合情报学、图书馆学与医学专业知识而形成的医学方法课程,其宗旨是增强医学生的信息意识,提高医学生驾驭知识信息的能力素养,掌握医学信息检索的技能。

（二）医学文献信息检索课的目标

作为医学方法学课程的医学文献信息检索，旨在培养学生的信息意识、信息能力和信息道德，掌握文献信息检索的基本理论、基本知识和基本技能，培养学生的检索、筛选、分析、评价、管理和综合利用文献信息的能力、创新能力和终身学习能力，在课程目标上必须注重两个强化，即核心概念与原理、关键技术与技能。对于理论层面的核心概念，不是单纯的重复性表述；对于基本原理，也不是抽象性阐述其内涵；对于实践层面的关键技术与操作，不是单纯的、反复性地重复演示、重复练习；而是将概念的内涵渗透在原理的阐述之中，原理的理论解析体现于技术的操作之中，在程序性的技术操作中深化对理论的理解，最终实现将知识转化形成技能。

具体体现在以下几个方面：

1. 培养学生的信息意识，提高学生对文献信息的敏感性，激发学生的信息需求和潜能。
2. 培养学生文献信息检索能力，使学生掌握文献信息资源的特点，灵活运用各种检索策略，高效地检索和获取文献信息。
3. 培养学生筛选、分析、评价文献信息的能力，对获取的文献信息能判断其权威性、可靠性和准确性。
4. 培养学生管理和综合利用文献信息的能力，进而激发学生的创新能力，达到终身学习的目的。
5. 培养学生自觉遵循学术规范、信息道德和信息法规。

（三）医学文献信息检索课的任务

医学文献信息检索课的主要任务是通过学习信息检索的基础知识和基本技能，了解医学及相关专业文献信息的基本知识，掌握常用医学信息资源数据库的使用方法，能够快速、准确地从各种医学信息资源中获取自己所需要的信息，并进行分析、评价和有效利用，从而使医学生的医学认知能力和医学创新能力得到提升。

（四）医学文献信息检索课的内容体系

医学文献信息检索课的内容由文献信息检索基础、文献信息检索语言、图书馆信息资源利用、中文数据库检索、外文文摘数据库检索、外文全文数据库检索、引文检索、专类信息检索、生物信息学数据库检索、网络信息资源检索以及文献信息分析与利用 11 个知识模块组成，它们相辅相成，共同构成了全面、系统、科学的教学内容体系。

（罗爱静）

第二节 信息素养及医学信息素养

一、信息素养的概念

诺贝尔奖获得者赫伯特·希曼 (Herbert Seaman) 指出：在过去，拥有知识意味着一个人必须把知识牢记，而现在，面临着纷至沓来的海量信息，拥有知识则意味着必须知道如何去获取有用信息。占有知识已经被掌握获取知识的通道所取代。美国学者路易斯·肖 (Luis Shaw) 认为：“懂得如何查找到所需要的知识，就等于掌握了人生全部知识的一半。其实很多知识留在待查状态就可以了，所以查询能力是一个成功人士的基本能力。”英国文豪塞·约翰逊 (S. Johnson) 也曾经说过：“知识有两种，一种是我们自己知道某主题的知识，而另一种则是我们知道什么地方能够找到知识。”因此，查找、利用信息的能力直接决定了解决问题的能力和学习生活的质量。

信息素养常被称作信息素质、信息文化、信息能力、资讯素养。1974 年，美国信息产业协会主席保罗·泽考斯基 (Paul Zurkowski) 在向美国图书情报学全国委员会提交的一份报告中将

其定义为:“所有经过训练的在工作中善于运用信息资源的人即称为具有信息素养的人,他们具有利用多种信息工具及主要信息资源使问题得到解答的技能。”随后美国图书馆协会(American Library Association, ALA)进行了十余年的研究和实践探索,于1989年发表总结报告,认为信息素养是信息时代的一种生存技能,具备信息素养的人能认识到何时需要信息,并能检索、评价和有效地利用所需的信息。具备信息素养的人是学会了如何学习的人,是能够进行终身学习的人。目前,信息素养较多地被定义为:从各种信息源中检索、评价和利用信息的能力,是信息社会公民必须掌握的终身技能。

二、信息素养教育及评价标准

(一) 国内外信息素养教育

1. 国外信息素养教育 信息素养教育起源于美国,其前称是图书馆利用教育。图书馆利用教育在20世纪70年代后半期进入到一个新的发展阶段,即信息素养教育阶段。从信息素养发展脉络分析,其教育大致分为三个阶段:

(1) 萌芽阶段(1980—1989年):标志性事件是1987年美国图书馆协会成立了信息素养教育委员会,目的是明确信息素养在学生学习和终身教育中成为一个良好公民过程中的作用,设计在正式或非正式学习环境下的信息素质教育模型,决定继续教育和教育培养的发展方向。

(2) 探索阶段(1990—1999年):主要是1990年国家信息素养论坛(The National Forum on Information)的成立和几所大学信息素养教育计划的制定。该论坛持续推动着信息素养教育在美国的发展。

(3) 发展阶段(2000年至今):这一阶段信息素养教育得到突飞猛进的发展,标志性事件是2000年1月美国大学与研究型图书馆协会(Association of College and Research Libraries, ACRL)通过了《高等教育信息素质能力标准》,促使美国大学开设了一系列信息素养教育课程。

目前国外信息素养教育采取了多样化的教学方式,其中以网络化教育为主,强调虚拟学习社区的构建和教学成员间的交互作用。许多高校和图书馆都设立了信息素养教育网站,如得克萨斯州立大学奥斯汀分校开发的基于Web的在线信息素质教育指南(TILT)。普渡大学图书馆与学院和研究图书馆协会联合开发了网络综合检索教育指南,为用户提供有关信息素养和信息检索方法等方面的课程内容。

2003年9月,联合国教科文组织(UNESCO)在布拉格召开的首届信息素养专家会议上发布了布拉格宣言:“走向信息素养社会”。宣言指出了信息素养是人的一种能力,是终身学习的一种基本人权。2005年11月,国际图联(IFLA)和UNESCO又在埃及发表了著名的《信息社会灯塔:关于信息素质和终身学习的亚历山大宣言》。表明信息素养教育已经成为一种国际化潮流,世界各国越来越重视信息素养的教育和推广。

2. 国内信息素养教育 我国高校的信息素养教育最早可以追溯到1981年《人民日报》刊载的刘毅夫、潘树广关于“建议在高校开设文献检索课”的文章;1984年教育部印发《关于在高等学校开设文献检索与利用课的意见》,规定高等学校都应该开设该课程,并突出图书馆在教学中的作用;1985年教育部下达《关于改进和发展文献课教学的几点意见》;1992年5月原国家教委再次印发了《文献检索课教学基本要求》,促使各高校着手制定教学大纲,为课程的蓬勃发展奠定了基础。2002年1月,首届“全国高校信息素质教育学术研讨会”在哈尔滨召开,标志着“文献检索课”向“信息素质教育”的转型。与会代表一致认为:在网络环境下,文献检索课发展的方向是向信息素养教育转变。在理论研究不断完善与丰富中,各高校纷纷进行信息检索课程的改革与创新。

(二) 国内外信息素养评价标准

1. 国外信息素养评价标准 自1989年起,美国大学与研究型图书馆协会(ACRL)就开始

为学术型图书馆拟定信息素养标准及指导方针。该协会于 2000 年批准并颁布的《高等教育信息素养能力标准》(*Information Literacy Competency Standards for Higher Education*)成为国际上公认的最具影响力的信息素养能力评估标准,该标准共分为 5 项一级指标、22 项执行指标和 87 项参考指标,其细则见表绪论 -1。

表绪论 -1 美国高等教育信息素养能力标准 (ACRL 标准)

标准一 能确定所需信息的性质和范围	能定义和说明信息需求 能确认各种类型和格式及潜在的信息源 能考虑获取所需信息的成本和效益 能重新评价所需信息的性质和范围
标准二 能高效地获取所需要的信息	能选用最适当的检索方法或检索系统获取所需的信息 能建构和实施有效的检索策略 能运用各种方法检索在线信息或个人信息 必要时能优化搜索策略 能写出摘要,记录和管理信息及其来源
标准三 能评判性地评价信息及其来源,并能把所遴选出的信息与原有的知识背景和评价系统结合起来	能概括所收集信息的中心思想和观点 能明晰信息的评价标准,并对信息及其来源进行评价 能综合主要思想与观点来创造新的概念 能比较新旧知识的差异与联系,确定新知识的含义与特点 能判断新知识对个人价值观的影响,并逐步调和冲突 通过与他人或者某一领域的专家、实践者对话,验证对信息的理解和解读是否正确 能确定是否需要修正原来的观点
标准四 能有效地利用信息达到某一特定的目的	能用新旧知识创造新的计划、新的作品及其他表现形式 能修正原先制定的工作 能与他人有效地交流
标准五 能理解与信息的使用和获取相关的经济学、法律和社会问题,并能够使信息的获取和使用符合伦理学和法律规范	能理解与信息和技术有关的许多伦理学、法律和社会经济学问题 能够遵守与信息资源获取和利用相关的法律法规、机构政策和礼仪 在进行产品和性能的交流时能够对信息源的使用表示致谢

上述标准衡量的是高校大学生通用层次的信息素养水平。近年来,为适应不同专业对信息素养的要求,ACRL 又陆续出台了一系列针对专业的信息素养标准。如 2006 年《科学与工程学科信息素养标准》,2010 年《心理学信息素养标准》,2011 年《新闻专业学生和专业人士信息素养标准》,2011 年《教师教育信息素养标准》以及 2013 年《护理学信息素养标准》等。

2. 国内信息素养评价标准 国内最早提出、内容比较完备并包含详细的二级评估指标的信息素养能力标准是陈文勇和杨晓光于 2000 年拟定的《高等院校学生信息素养能力标准》;北京高校图书馆学会于 2005 年提出的《北京地区高校信息素质能力指标体系》,是我国第一个比较完整、系统的地区性信息素养能力标准评价体系;2008 年 4 月,北京地区高校图书馆专家提出了《高校大学生信息素质指标体系》(讨论稿),为高校实施信息素养教育和评价人才综合素质提供重要指标和依据,并对信息素养教育的研究起到很强的指导作用。但国内至今未出台适合我国国情的信息素养评价指标体系的正式文件与权威性的国家标准。

三、医学信息素养内涵及发展

(一) 全球医学教育最低基本要求与医学信息素养

医学是一个对信息依赖性极高的学科,要想成为一名合格的医生,具备信息素养尤为重要。

1999年6月,纽约中华医学基金会(CMB)理事会批准资助,成立了国际医学教育专门委员会(IIME)。其任务是为制订“全球医学教育最低基本要求”(以下简称“基本要求”)提供指导。“基本要求”是指世界各地医学院校培养的医生都必须具备的基本素质,包括医学知识、临床技能、职业态度、行为和职业道德等四大方面。将这些领域所需的能力与信息素养内涵相对照可以看到,信息素养几乎渗透进医学教育的所有领域。

1. 涉及信息意识和信息道德的条目 “自我调整的能力,认识到不断进行自我完善的重要性和个人的知识和能力的局限性,包括个人医学知识的不足等”,“认识有关患者文件、知识产权的权益、保密和剽窃的伦理和医学问题”。

2. 涉及信息能力的条目 “注意倾听、收集和综合与各种问题有关的信息,并能理解其实质内容”“建立和妥善保管医疗档案”“能综合并向听众介绍适合他们需要的信息,与他们讨论关于解决个人和社会重要问题的可达到的和可以接受的行动计划”。

3. 涉及信息分析与利用的条目 “采集包括职业卫生等在内的相应病史资料”“运用基本的诊断和技术规程,对获得的观察结果进行分析和解释,确定问题的性质”“运用循证医学的原则,在挽救生命的过程中采用恰当的诊断和治疗手段”“进行临床思维,确立诊断和制订治疗方案”“发展独立、自我引导学习的能力,以便在整个职业生涯中更好地获取新知识和技能”。

4. 涉及信息管理全过程的条目 “从不同的数据库和数据源中检索、收集、组织和分析有关卫生和生物医学信息”“从临床医学数据库中检索特定患者的信息”“运用信息和通讯技术帮助诊断、治疗和预防,以及对健康状况的调查和监控”“懂得信息技术的运用及其局限性”“保存医疗工作的记录,以便于进行分析和改进”“建立和妥善保管医疗档案”。

5. 涉及信息评价和批判性思维的条目 “在职业活动中表现出有分析批判的精神、有根据的怀疑、创造精神和对事物进行研究的态度”“懂得根据从不同信息源获得的信息在确定疾病的病因、治疗和预防中进行科学思维的重要性和局限性”“应用个人判断来分析和评论问题,主动寻求信息而不是等待别人提供信息”“根据从不同来源获得的相关信息,运用科学思维去识别、阐明和解决患者的问题”。

由此可见,信息素养在国际医学教育中具有相当重要的地位。同样,我国也对医学教育中的信息素养教育提出了要求,2008年11月,教育部、卫生部颁发的《本科医学教育标准——临床医学专业(试行)》,该标准包括毕业生应达到的基本要求和医学本科教育办学标准两个方面的内容。毕业生应达到的基本要求按照思想道德与职业素质、知识目标和技能目标三个方面逐一列出。其中,思想道德与职业素质项,第5条“树立终身学习观念,认识到持续自我完善的重要性,不断追求卓越”;技能目标项,第10条“结合临床实际,能够独立利用图书馆和现代信息技术研究医学问题及获取新知识与相关信息”,第12条“具有自主学习和终身学习的能力”。

医学本科教育办学标准的10个领域的“教育资源”项中,“图书及信息服务”要求“医学院校必须拥有并维护良好的图书馆和网络信息设施,必须建立相应的政策和制度,使现代信息和通讯技术能有效地用于教学,使师生能够利用信息和通讯技术进行自学,获得信息,治疗和管理患者及开展卫生保健工作”。该项的注释为“学校必须高度重视图书馆的建设和投入,每年图书文献资料购置经费及其占学校当年教育事业经费拨款的比例必须达到国家有关规定的要求”。

(二) 医学信息素养的内涵

信息素养内涵应该包括信息意识、信息能力和信息道德三个方面。

1. 信息意识 就是主观上对信息具有强烈的知情愿望和高度敏感,具体又可划分为三个方面:领域意识、前沿意识和线索意识。领域意识是指医学生对其所从事的学科或专业领域信息的关注程度;前沿意识是对学科或专业领域及其相关学科或专业领域发展前沿的关注程度;线索意识是对学科或专业领域的再现事件保持记忆、及时关联和发现线索的敏感度。



2. 信息能力 是信息素养的核心,培养医学生的信息能力至关重要。它又体现在以下几方面:

(1) 信息获取能力:指医学生参与各类活动获取信息的能力。包括文献型信息获取能力、实物型信息获取能力和语言型信息获取能力。

(2) 信息积累能力:指医学生自觉地、有目的地累积、储备信息的能力。应避免“信息积累就是收集和物理储存了大量相关信息”的认识误区,力求在物理储存相关信息的同时,将这些信息深刻地记忆在脑海中,并能够在需要时随时提取。

(3) 信息表达能力:包括语言表达能力和文字表达能力。医学生不仅要能够通过课堂讨论、参加学术会议等方式口头交流传播信息,也要能够以书面形式发表论文,能让大家一起讨论共同关注的话题。

(4) 信息甄别能力:甄别是一项极其复杂的工作,往往需要分析该信息是在什么背景下出现的,由什么人发布的,在什么场合发布的,通过什么途径传播的等。包括信息的可靠性、可信度和可能性。信息可靠性甄别的是信息“有没有”,一般情况下来源于政府报告、备忘录、听证会报告、政府文件、官方网站、机构通报、学术期刊、科学文献等的信息可靠性较大。信息可信度是指信息的性质是否可信,甄别的是信息“真不真”。信息可能性是指信息内容是否可能,甄别的是信息“对不对”,这就需要医学生具备该学科的坚实基础。

(5) 信息利用能力:信息利用是上述能力的最终目的。医学生信息利用能力主要表现在:利用获取的信息提高和弥补个人知识和认识的不足;利用信息技术建立个人信息资料库;利用获取的信息和积累的知识重构个人的知识体系;利用从不同来源获得的信息进行临床医疗决策,解决患者的实际问题;利用医学信息进行医学科研课题研究,对医疗成果进行总结、转化和推广等。

3. 信息道德 是指个体在信息活动中应遵守的道德规范,简单地讲就是要合理合法地加工、传播和利用信息。对于医学生来说,一方面要了解与信息使用相关的经济、法律和社会因素,比如网络信息安全、知识产权保护等问题;另一方面获取和利用信息要符合法律和道德规范的要求,比如不合理使用图书馆文献,学术论文中引用他人的文献应注明,利用患者信息时要保护患者隐私等。

信息意识、信息能力和信息道德三者相辅相成,互为补充。信息意识是前提,对信息素养起着决定性作用,控制着信息行为的发生。信息能力是核心,加强信息能力,才能更好地掌握信息检索技能,提高查找信息的效率,使信息创造力得到更大的发挥,本书重点介绍的是信息能力部分。信息道德是保障,它保证个体的信息行为遵循学术规范和社会共识,从而维护信息社会的正常秩序。必须承认,良好的信息意识和信息道德的培养是一个潜移默化的过程,需要经过长期教育和个体主动认同后才能得以内化。提高信息能力可以有力地强化信息意识,而增强信息意识又可以更有效地促进信息能力的提升。

(三) 医学信息素养的发展

医学领域的信息素养研究始于1975年Hart L发表于《美国药学教育杂志》(*American Journal of Pharmaceutical Education*)的“药物信息技能的教学方法”一文。时隔十年之后(1985年),“information literacy”一词才首次正式出现在与医学信息检索有关的文献中。自2001年《全球医学教育最低基本要求》发布以来,其对医学生基本能力中信息素养能力的强调使得该领域的研究文献呈现迅猛增长的态势。

新世纪伊始,各国已经开始重视医学领域的信息素养并将其作为政府行为加以实施。2000年,由加拿大多伦多大学卫生与健康领域的相关院系共同参与建立了世界上第一个较为成熟和规范的知識转化项目(knowledge translation program),强调将循证医学证据有效地转化于临床实践。由此提出知识转化的基本概念,即有效、及时地将循证医学证据应用于卫生保健实践,

促进高质量医学信息的获取和应用,以获得卫生保健的最佳效果。2003年,美国医学图书馆学会(Medical Library Association, MLA)成立了一个健康信息素养任务组,旨在提高卫生保健专业人员的信息素养。2006年 MLA 主席 Jean P. Shipman 将优先发展领域确定为 MLA 成员在提供面向公众的健康信息方面要发挥更大的作用,由此开始将信息素养的教育对象由卫生保健提供者拓展至整个公众。

2006—2008年 MLA 受美国国立医学图书馆(National Library of Medicine, NLM)委托执行一项健康信息素养(health information literacy, HIL)研究计划,该计划的总体目标是增加卫生保健提供者对健康信息素养的了解,证明医学图书馆员能够在处理健康信息素养问题中起作用,并增加卫生保健提供者和消费者对 NLM 卫生信息工具的认知和利用。作为一项长期计划, HIL 计划开发的课程及相关的继续教育资源可通过 MLA 网站的健康信息素养页面获得。最新的两个卫生信息素养教程“信息处方:解决健康信息素养问题”和“信息处方:图书馆员用来帮助卫生专业人员提高健康信息素养的工具”,可链接到 MLA 最新健康信息素养继续教育课程列表。

2008年,英国国家卫生服务系统(NHS)在苏格兰地区实施的一个教育草案中,提出了“优质信息促进优质卫生保健(better informed for better health and better care)”的理念,并建立了一个促进苏格兰地区卫生保健的信息素养框架。苏格兰政府的行动计划“Better Health, Better Care”认识到充分利用信息以建立一个更加健康的苏格兰的必要性,并要求“卫生和社会保健专业人员需要学习快速获取信息并能合理利用信息的技能”。不仅如此, NHS 还试图创建一个国家卫生信息支持服务系统,以确保患者和公众在查找所需信息、理解获取信息、并培养有效利用信息的知识和技能等方面获得支持。信息和知识是用来改善患者保健和促进公众健康的有力工具,信息素养现已成为个人或社区促进健康、幸福以及改善生活环境的关键因素。NHS 还创立了“将信息学嵌入临床教学计划(embedding informatics in clinical education, eICE)”,通过网站提供在线学习与培训工具,对医务人员进行信息学教育。2012年, eICE 出版了课程导航《学习管理健康信息(2012)》(*Learning to Manage Health Information*),帮助临床教育者把信息学内容整合入临床教育。

本章小结

本章在分析医学人才的知识能力结构要求的基础上,探讨了医学文献信息检索课程的作用、目标与任务,阐述了信息素养、医学信息素养的内涵、标准和发展。

(张士靖)

思考题

1. 何谓“global, minimum essential requirement medical education”?
2. 试分析医学文献信息检索课程在高等医学人才的综合素质培养中的作用。
3. 何谓创新型医学人才?你认为医学文献信息检索课程将对你成为创新型医学人才起到什么样的作用?
4. 医学文献信息检索课程的性质、目标和任务是什么?
5. 何谓信息素养?你认为培养自己的信息素养对你的学习和工作将产生怎样的影响?
6. 作为医学生,其信息素养要求与普通高校的大学生有哪些不同?
7. 学习本章后,你认为应该怎样提高自己的信息意识?
8. 美国高等教育信息素养能力标准主要包含哪些一级指标和二级指标?以该标准衡量自己,你认为自己的信息素养如何?

9. 在平时的专业课学习中,授课教师是否注意培养你们的信息素养?你认为应该如何加强专业教师与信息素养教育的联系?

10. 通过本章的学习,你对信息检索课程的重要性从认识有了哪些提高?请提出你的学习目标和计划。



第一章 文献信息检索基础

文献,作为人类所特有的承载和传递文明的载体,记录着人类在漫长的历史长河中逐渐积累的经验 and 创造的知识,成为人类不断发展进步的智慧源泉。进入高速发展的信息时代,信息和知识,演化为这个时代最具特征性的两个概念。那么,什么是文献、什么是信息、什么是知识、它们存在的状态和获取的方式如何等,这些知识内容成为医学专业人员充分认识并有效利用现有文献信息资源的基础和前提。

第一节 文献信息检索概述

犹如医学知识体系的构建,文献信息检索领域同样是由许多经典概念及其相互联系编织起来的知识框架。

一、文献信息概述

(一) 文献信息的定义

1. 文献(literature, document) 国家标准局于1983年颁布的《文献著录总则》(GB3792.1-83)中将文献一词定义为:“文献是记录有知识的一切载体。”在这看似简单的定义中,实际上包含了作为文献的四个基本要素:

- (1) 记录知识的具体内容。
- (2) 记录知识的手段,如文字、图像、符号、声频、视频等。
- (3) 记录知识的物质载体,如纸张、光盘等。
- (4) 记录知识的表现形态,如书、刊,以及电子版等多种形态。

由此可见,人类积累创造的知识,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,都称为文献。记录科技知识的文献即称为科技文献。

在查找和阅读文献时,人们关注的不仅是文献的载体和形态,更注重文献中所传递的信息和蕴涵的知识。文献因载有知识和信息才有其存在的价值和意义,而知识和信息因附着于文献这一载体之上,才得以超越时空地保存和传递。

2. 知识(knowledge) 是人们在认识和改造客观世界的实践中所获得的认识和经验的总和。是人类通过对信息的感知、获取、选择、处理、加工等一系列思维活动和过程,形成的对客观事物的本质和规律的认识。运用知识,实际上就是遵循了事物的客观规律,这是进一步认识和改造客观世界的最基本前提和基础,这也正是知识经济这一理论的逻辑起点。因为从本质上讲,知识蕴涵着推动社会发展和人类进步的无穷力量。

从信息论的角度看,知识是人类大脑中重新组合形成的序列化信息。往往是为了实现某种特定目的,在同种或相关信息积累的基础上,经过抽象化了的的信息。所以,属于意识范畴的知识,虽然是思维的结果、人脑的产物,却以信息为原料、以信息的获取为前提、以信息的利用为核心。

3. 信息(information) 国家标准《情报与文献工作词汇基本术语》(GB4894-85)中对信息的定义为:“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定性减少。”既然是事物的一种普遍属性,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形态及特征等,如风、雨、雷、电,春、夏、秋、冬;而社会信息反映人类社会各种事物和现象的状态及性质,如语言、战争、贫困,人的喜、怒、哀、乐等。在科学技术领域,人们将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,例如:科学家破译的依附在人类体细胞的细胞核中染色体基因上的信息。

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。人类正是通过对丰富多彩的信息的捕捉、感受、提炼、加工,才一点一滴地逐渐认识了这个奇妙的大千世界。之所以如此,源于信息所具有的基本属性:

(1) 客观性:物质的客观存在即物质本身的客观性,决定了表征其存在的信息的客观性。

(2) 可知性:信息作为客观事物的反映,可被人类通过不同方式、不同方法、不同手段所感知和认识。因人类认识能力所限,只有尚未被感知和认识的信息内容,而不存在不能被感知和认识的信息。正因为如此,信息也才成为人类认识世界的有效手段。

(3) 传递性:经人类感知(捕捉、获取)的信息,可以各种方式、借助于各种载体或媒介广泛传递,为更多的人所接收和利用。

综上所述,通过对表征客观世界之客观信息的获取、加工等一系列思维过程,形成了反映客观事物本质和规律的具有主观性色彩的知识,将知识以某种方式系统化地记录于某种载体之上,便形成了具有一定表现形态的文献。三者的逻辑关系可分别用图 1-1 表示其形成的递进关系,图 1-2 表示其外延的相容关系。

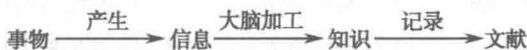


图 1-1 信息、知识、文献的形成

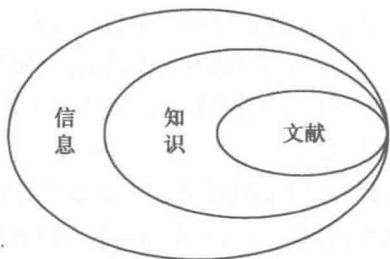


图 1-2 信息、知识、文献的相容关系

由此,进一步解读“信息资源”的含义,是指以各种记录形式存在的信息载体,其外延大于文献资源。但基于有用的知识信息多以文献的形式存在这一客观事实,所以从用户的角度可不必严格区分信息资源与文献资源这两个概念,也可用“文献信息资源”一词涵盖这两个概念。资源,一般指人类生存与发展所必需的物质基础。因此,文献信息才被视为支撑现代社会文明进程的三大支柱之一,成为促进知识经济发展最重要的智力资源。

(二) 文献信息的类型

人们认识事物往往是从区分事物开始的,即比较和分类。而比较和分类之前提就是标准(分析事物的角度或维度)的选择与确立。对文献信息的认识及类型的划分亦不例外,这种认识和划分的目的是借此深化对文献信息本质及规律的理解和认识。根据不同的划分标准,可将文献区分为不同的类型。

1. 按载体的不同,常将文献划分为如下几种类型:

(1) 印刷型:以纸张为载体,以印刷技术为记录手段而产生的文献类型,如传统的图书、期

刊等。便于直接阅读,符合传统阅读习惯,因此成为人们信息交流和知识传递的最重要、最常用媒介。但缺点是存储密度小、占用空间大,易受虫蛀、水蚀,不宜长期保存。

(2) 电子型:采用电子手段,将文献信息数字化(通过编码和程序设计将文献的原有语言形式变成计算机可存取、阅读的数字化形式),储存于磁盘、光盘等载体上,并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。主要包括以数字化的方式将文本、图像、声音等多媒体形式表述的知识信息存储于光和磁介质上,借助于网络传播并通过计算机等设备进行阅读或下载的连续性出版物(称为电子期刊或数字化期刊)或书籍(称为电子图书或数字化图书),及其信息集合(称为全文数据库)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,开辟了一种新的信息传播渠道,极大地提高了文献信息的传递效率,加快了社会信息化的进程。目前,电子型文献信息正以其容量大、形式多、出版快、成本低,以及检索、阅读、复制便捷等独特的优点为越来越多的人所接受和利用。

(3) 声像型:又称声像或直感资料。指以磁性或感光材料为载体,以特殊方式直接记录声音和图像所产生的一种文献形式,包括唱片、录音带、录像带,以及高密度视听光盘等。既可长期保存,又可反复播放和复制。使人们通过形象、逼真、生动的动态信息更为直观地感受和认识事物。声像型文献信息通过数字化转换成多媒体形式已成趋势。同时,以传统感光材料为载体,用摄影方法制作的胶卷或胶片等缩微型文献亦多为数字化方式所取代。

2. 按出版类型,又常将文献划分为如下几种类型:

(1) 图书:是文献中最为古老的,至今仍被频繁使用着的一种文献类型。也是传统图书馆最主要的馆藏内容,图书馆亦因收藏图书而得名。除了记载有知识信息这一本质特征外,联合国教科文组织(UNESCO)对篇幅(封面除外)不少于49页的非定期出版物称为图书,以示与期刊等连续出版物的区别。虽然图书的种类繁杂、形式多样、功能各异,但就学习与研究而言,常用的图书有教科书、专著、参考工具书等。

在每一种正式出版的图书的版权页或其他明显部位都标有一个由10位数字组成的国际标准书号(international standard book number, ISBN),形式如:ISBN 7-5601-2748-7/R·21。这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,读者可借此通过某些文献信息系统查询某种特定图书。

(2) 期刊:是指具有相对固定的刊名、相对固定的编辑出版单位、相对固定的出版周期、相对固定的报道范围,旨在以分期形式报道最新知识信息且逐次刊行的连续出版物。以报道最新科技知识、揭示最新科研成果为主的即为科技期刊。科技期刊主要刊载属一次文献范畴的科技论文,因而成为科研人员展示成果的园地和实现知识更新的知识源泉。

学术性期刊的刊名中常有“杂志”(journal)“学报”(acta, bulletin)“纪事”(annals)“文献”(archives)“会刊”(proceedings)“评论”(reviews)“进展”(progress, advances in...)等字样。

同图书的ISBN一样,每种期刊均有一个由8位数字组成的国际标准连续出版物号(international standard serial number, ISSN),例如:ISSN 0317-8471。ISSN同样具有唯一性和专指性,因而成为读者查询某种刊物的一个检索途径。ISBN和ISSN常被用做区分正式或非正式出版物的判断标准。

(3) 会议文献:学术会议是进行学术交流的一种重要方式和渠道。在世界范围内每年召开的医学各学科专业的学术会议甚多,产出的会议论文数以万计。会议文献,有的在会前出版预印本,有的在会后结集出版会议文集,还有的是将会议文献压缩为500~800字的大摘要,结集以增刊或专集形式刊发于各种学(协)会的机关刊物上(国外期刊中常见),只有很少一部分会议文献日后能在各种学术期刊上正式发表。由于会议文献表述的往往是科研人员的最新成果,尤其是一些阶段性成果,与会者不仅可以借此展示自己的成果,更可以通过互相交流,获取许

多有价值的信息和有益的启示,因此,倍受专业人员的青睐。

(4) 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。

可见,学位论文是学生学习成果、学习水平、学习能力的一种标志和体现,只不过这种学习强调的是研究性质的学习,是以自己独立获取知识、掌握知识、运用知识,乃至创造知识为特点。

(5) 科技报告:是科学技术报告的简称。根据国家标准(GB7713-87)给出的定义: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试验和评价的结果;或是论述一项科学技术问题的现状和发展的文件。

科技报告是为了呈送给科学技术工作主管机构或科学基金会等组织或研究的主持者等。科技报告中一般应该提供系统的或按工作进程的充分信息,可以包括正反两方面的结果和经验,以便有关人员和读者判断和评价,以及对报告中的结论和建议提出修正意见。可见,科技报告旨在提供系统、详实的信息,不以发表为目的,是科研历程及其成果的完整记载,因而可具保密性。

(6) 专利文献:专利(patent),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件的总称,是受专利法保护的有关技术发明的法律文件。专利文献记载着发明创造的详细内容及被保护的技术范围的各种说明书(亦称专利说明书),是集技术、法律、经济信息于一体的特殊类型的科技文献。虽然绝大多数的医学研究成果(如揭示某种疾病的本质)属于科学发现的范畴,而不属于发明创造的范围,或虽然是一种新的技术方法(如实验方法、诊疗技术),但因不具备工业实用性,因而均未能享有专利保护,但涉及的诸如各种医疗设备、仪器、药品、化学物质、微生物菌种等却都属于专利保护的范畴。因此,专利文献对医学研究和医疗工作同样有着重要的参考价值。

(7) WHO 出版物:作为统筹、协调全球公共卫生事业的世界卫生组织(WHO),经常围绕全球公共卫生的重大问题,或地区性的特殊事件等,以学术文件的形式发布信息通报、传播科学知识、交流工作经验等,形成一类具有独特学术价值的 WHO 出版物。这类出版物主要涉及两大系列:丛书和期刊。丛书为不定期出版,每种书作为一个独立的单元——专著单独出版。如《技术报告丛书》的每一种专著都是 WHO 专家委员会就某一特定卫生或医疗问题推荐的通用标准、指南或研究报告。WHO 出版的期刊,如 *Bulletin of WHO* (《世界卫生组织通报》)、*International Digest of Health Legislation* (《国际卫生法规文摘》)、*WHO Drug Information* (《WHO 药物信息》)、*Weekly Epidemiological Record* (《疫情周报》)等。其中刊载的文献多为国际著名的生物医学文献信息系统所收录,如美国的 PubMed、荷兰的 Embase 等。可通过访问 <http://www.who.int/dsa> 了解 WHO 出版物的详细内容。

(三) 文献信息的级别

根据文献中信息含量的多少、内容加工深度的差别,以及功能作用的不同,常将其划分为以下四个级别:

1. 一次文献(primary document) 是指作者以其本人的研究成果(如实验、观察、调查研究等的结果)为基本素材写成的原始创作,如专著、期刊论文、科技报告、学位论文等。

一次文献,所记录的是作者的最新发现或发明,以及新的见解、新的理论、新的方法等新颖、具体而详尽的知识,因而成为科学研究等工作的最主要信息来源,尤其是期刊论文,已成为科技文献的主体。但由于其量大、分散而无序,给读者的查找与利用带来极大的不便。

2. 二次文献(secondary document) 是将大量无序、分散的一次文献收集、整理、加工、著录其著者、篇名、分类、主题、出处等特征,并按一定的顺序加以编排,形成供读者检索所需一次

文献线索的新的文献形式,包括索引、文摘、目录及相应的数据库(参见本章第二节)。因其具有的检索功能而被称为检索工具或检索系统。

从上述定义的引申来看,二次文献信息是关于文献的文献、关于信息的信息。因此,现在网上的 Yahoo、Google 等主题指南、搜索引擎,这些关于数据库和网页的信息集合,其功能作用等同于上述二次文献,所以称其为网络检索工具,指提供网上信息资源检索和导航服务的专门站点或服务器。它同样是对采集到的网上信息进行加工整理,建立起存储和管理网络信息的索引数据库,为用户提供网络信息检索导引。

相对于一次文献而言,二次文献是从分散到集中、从无序到有序、从繁杂到简约,因而具备了可查检的便捷性,用以解决读者查阅所需特定文献线索的问题。知识和信息的海量无限性与人们时间、精力的相对有限性,知识和信息的散乱无序性与用户使用的特定选择性之间的矛盾,一直是困扰学者学术生涯的两道永恒的难题。在知识爆炸、信息泛滥的今天,这两个矛盾愈加突出,仅靠“学海无涯苦作舟”的勤奋和坚韧,实难以登上光辉的彼岸,只有驾驭好二次文献这一叶方舟,方可自由徜徉于知识和信息的海洋。正因为如此,包括网上检索工具在内的二次文献及其利用也就成为信息检索课的核心内容。

3. 三次文献(tertiary document) 是科技人员围绕某一专题,借助于二次文献,在充分研究与利用大量一次文献的基础上,即经过阅读、分析、归纳、概括,撰写而成的新的文献,或综述已取得的成果进展,或加评论、或预测发展趋势。形式如综述(review)、述评(comment)、进展(advance; progress)、现状(update)、发展趋势(trend)等期刊文献和百科全书、年鉴、手册等参考工具书。许多学术期刊上均辟有综述栏目,而且出版专门刊载三次文献的综述性期刊亦越来越多。

与一次文献的产生所不同的是,三次文献是以现有一次文献中的知识信息为基本研究素材,对其进一步的加工、整理、重组,使之成为更加有序化的知识信息产品。但由于同样融入了作者的智力劳动,因此,同一次文献一样同属智力产品。三次文献具有信息含量大、综合性强和参考价值大等特点。可使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、某一领域当前的研究水平、动态。

4. 零次文献(zero document) 指未经信息加工,直接记录在载体上的原始信息,如实验数据、观测记录、调查材料等。这些未融入正式交流渠道的信息,往往反映的是研究工作取得的最新发现,或是遇到的最新问题,或是针对某些问题的最新想法等,而这一切无疑是启发科研人员的思路、形成创造性思维的最佳思维素材。

此外,学术界还常将通过非正常交流渠道获得的、非正式出版物称作灰色文献(grey literature)。灰色文献和零次文献的概念内涵虽有一定程度的重叠,但作为一般的专业人员可不严格区分这两个概念。

综上所述,各级别文献的形成及其相互关系,可用图 1-3 表示。

图 1-3 不仅展示了各级别文献递进的逻辑关系,而且体现了信息工序化的进程,更从信息学角度揭示了科学研究的本质所在。这一综合性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) 文献信息的特点

随着科技发展与人类文明进程的日益加快,科技文献信息亦呈现出相应的变化,主要表现在如下几个方面:

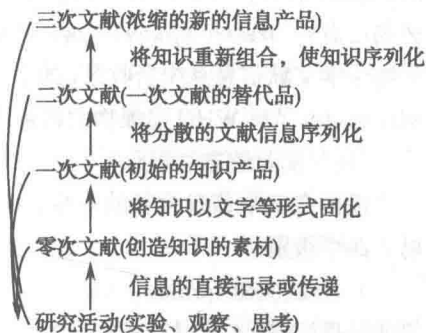


图 1-3 各级别文献的形成



1. 数量庞大, 增长迅速 据报道, 全世界的科技期刊数量早已超过 10 万种, 每年发表的论文数量超过 400 万篇。我国的科技期刊已超过 5000 种, 年发表论文已超过 20 万篇。无论是期刊的数量, 还是论文的数量, 每年均以超过 10% 的增长速度递增。就医学领域而言, 常用的外生物医学期刊近 3000 种, 国内出版的生物医学期刊超过 1000 种。

2. 文种繁多, 但呈明显的英文化趋势 第二次世界大战前后, 世界科技发展中心首次移到美国, 英文开始取代德文和法文而成为一种国际性科技交流的语言。因此, 尽管国际上科技文献的文种繁多, 但英文所占的比重正呈直线上升。尤其是互联网的广泛运用, 更加速了这种英文化的趋势。许多非英文国家为顺应这一发展趋势, 纷纷作出积极的反应。如日本出版的英文医学期刊就多达 150 种, 即使是在日文医学文献中, 英文专业词汇、英文图表亦比比皆是。我国出版的英文期刊的数量也在迅速增加, 而且许多学术期刊, 如《中华外科杂志》《中国循证医学杂志》等均以优先发表作为鼓励, 来吸纳英文文稿。

3. 内容交叉重复 现代科学技术的发展呈现两种趋势, 一是学科的分化和专深化, 二是学科的交叉融合、综合化。前者导致学科的划分越来越细、越来越专, 后者导致许多交叉学科和边缘学科的产生。这种趋势必然导致科技文献内容上的交叉、重复。

4. 文献形式呈现多样化 随着现代社会交流渠道的增多和交流方式的变换, 科技文献信息的交流与传递形式亦呈现多样化, 表现在: ①文献以一种形式发表后, 又以其他形式重复发表, 如会议论文与期刊论文; ②文献以一种文字发表后, 又以另一种文字重复发表, 如中文与英文; ③文献以一种载体发表后, 又以另一种不同载体出现, 如印刷本期刊、文献数据库、网络期刊等。

5. 文献分布既集中又分散, 但呈专题化或专集化趋势 就医学领域而言, 由于学科专业的专业化、综合化发展, 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。加之文献的数量和期刊的品种骤增等因素, 就使得与某一专题有关的文献往往分散在众多期刊上, 给读者的特定需求造成不便。

为此, 许多医学期刊采取专题化或专集化形式出版, 即在期刊的某一期以数篇、甚至数十篇文献集中讨论一个专题。在这方面最为突出的应属日本出版的医学期刊, 不仅综合性期刊如《日本临床》, 专科性期刊如《内科》, 甚至像《高血压》等以一种疾病命名的专业性很强的期刊, 亦多采用专集化形式出版。

6. 知识老化加快, 文献寿命(半衰期)缩短 由于科学技术发展的速度越来越快, 新知识的产生也日益加快和增多, 致使已有知识越来越快地被新知识所淘汰, 文献的使用寿命也必然随之缩短。由此, 人脑中已掌握的知识也不得不随之不断更新以跟上时代发展的步伐。这便是倡导继续教育、终身学习以及学习型社会的根本原因。

7. 交流传播及变化速度加快 以往以书信、期刊论文等形式互通信息, 进行学术交流。如今, 论文的编辑、印刷、出版、发行等一系列环节都可以借助于网络, 无论是发 E-mail, 还是查阅电子书刊, 信息的传播已是瞬息之间的举手之劳, 偌大的世界宛如小小的地球村。以往单纯的印刷型书刊, 在短短的几十年间派生出缩微型、视听型, 以及光盘、网络等多种形态, 这种变化既是信息时代的产物, 更是信息时代的标志, 极大地丰富了现代人对文献信息的多样化需求。

(五) 文献信息的社会功能

1. 科技文献是科学的表现形式, 是汇集人类财富的主要载体 自从有了文字以来, 人类在生存与发展的漫长历史长河中, 将创造的知识, 获取的经验, 常以文字的形式记录保存下来, 科学知识更是如此。因此, 科技文献也就成了人类的精神财富和世代文明之结晶。英国哲学家培根的名言: “知识就是力量”, 邓小平同志提出“科学技术是第一生产力”, 都揭示了这样一个真谛, 那就是科学知识、科学技术能够推动生产力的发展, 进而为人类创造更多的财富。知识创新是以充分挖掘、开发、利用现有知识资源宝藏为前提。

2. 科技文献是衡量科学技术水平的重要依据 科技文献是记录、揭示、传播最新科研成果和科学知识的重要手段和媒介。因此,科技文献产出的数量和质量往往成为衡量一个国家或地区,一个学科或一个单位科技发展水平乃至某位学者的学识水平的标志,是体现其科学创造力和确认其学术地位的一个公认的指标。用以进行文献学研究的文献计量学,揭示科学发展时空特征及规律的科学计量学,探究信息传播与利用之规律和特征的信息计量学,乃至科学评价的科学研究,无一不以科技文献作为研究的最直接素材。离开科技文献,这些研究工作可谓难为无米之炊。

3. 科技文献是传播科学知识的最基本、最主要的手段 尽管人们可以通过书信、各种学术会议等形式进行学术交流,但最重要、最基本,也是最为规范的手段和途径只有科技文献,尤其是期刊论文。它不仅是科技人员展现、揭示、报道自己的科研成果的园地,同时也是科技人员获取最新知识、实现知识更新的最主要信息来源。

4. 科技文献是确认科学发现与技术发明优先权的认证依据 学术界有这样一个惯例,某项发现、发明的大部分荣誉,倾向于由第一个发表这个观点、学说或第一个发表这个发现的人获得,而不是由第一个发现它的人获得。科学发展史上的任何一项发现或发明总是和第一个发表相应文献的作者姓名联系在一起,如血液循环学说与威廉·哈维、细菌的发现与巴斯德。也就是说,科技文献成为认定科学发现与技术发明优先权的重要凭证。科技人员,在获得新知识、新发现等研究成果时,都要尽快、尽早以论文形式发表(或申报专利),以期获得社会的承认。

二、文献信息检索概述

(一) 文献信息检索的定义

所谓的文献信息检索,是指在一定信息需求驱使下,利用现有文献信息资源有效获取所需文献信息内容的活动及其过程。

文献信息检索的内涵有广义和狭义之分,广义的包括文献信息的存储(storage)和查找(retrieval)两个过程。文献信息的存储,是指将大量无序的文献信息集中起来,根据其内容特征和形式特征,进行分类、标引、浓缩等一系列的加工处理,并按照一定的工作规范和技术要求将其存储于一定的载体或介质之中,形成具有检索功能的有序化信息集合,表现为检索工具或检索系统。文献信息的查找,是指运用编制好的检索工具或检索系统,获取满足用户信息需求之文献信息的活动。狭义的文献信息检索,往往仅限于后者,即利用信息检索系统获取特定的文献信息。

(二) 文献信息检索的类型

1. 根据检索手段的不同,将文献信息检索分为手工检索、联机检索、光盘检索和网络检索。

(1) 手工检索:利用目录、文摘、索引、题录等印刷型或纸质版检索工具查找和获取文献信息。

(2) 联机检索:是指用户借助数字通讯网络,在本地终端设备上,对远程联机检索中心的数据库进行直接的人机对话式查找。目前,许多大型的国际联机检索服务系统如 DIALOG、STN (the Scientific & Technical Information Network) 等不断推出基于 Internet 平台和 Internet 的网络产品。

(3) 光盘检索:将文献信息记录在光盘(compact disc, CD)这种新型的存储载体上,制成光盘数据库。借助光盘检索系统,可以不受地域与时空的限制,查找和利用其中的文献信息资源。随着硬盘技术和网络化信息共享技术的不断发展,将光盘数据库的内容镜像或缓存到硬盘中,然后通过本地局域网提供光盘检索服务。

(4) 网络检索:此处所述的网络检索是一个狭义的概念,是相对于联机检索和光盘检索而言的,主要指互联网信息检索。目前,网络检索的主要方式是搜索引擎,它是互联网的一种信息检索工具,包括信息收集、存储管理和提供检索。自 1994 年问世以来,成为人们获取网络信



息的主要工具之一。无论是联机检索还是光盘检索,都日益融入网络检索当中。

2. 根据数据格式和检索技术的类型不同,将文献信息检索分为:文本信息检索、多媒体信息检索、超媒体及超文本信息检索。

(1) 文本信息检索:文本(text),是数字化资源中最常见的形式,主要包括二次文献数据库和全文数据库,前者仅能检索文献的线索(即题录)和文摘,而后者是将文献全文的全部内容转化为计算机可以识别和处理的信息单元从而形成的数据集合,并进行全文本的词(字)、句、段落等深层次的编辑、加工,以及标引、抽词、排序、索引编制。因而全文检索可以直接根据文献资料的内容进行检索,支持多角度、多侧面的信息资源的综合利用。

(2) 多媒体信息检索:多媒体技术是把文字、声音、图像(形)等多种形态的信息,通过计算机进行数字化加工处理而形成的一种综合性信息传播技术。多媒体信息检索就是以多媒体信息为检索对象的信息检索,包括视频检索、音频检索、图像检索等。

(3) 超媒体及超文本信息检索:传统的文本是一种线性结构,用户必须顺序阅览。而超媒体是一种非线性的网状结构,用户可以沿着交叉链接选择自己感兴趣的部分阅读。超文本早期多为文字信息,现在已经容纳包括图像(形)、视频、声频等各种动态、静态信息。这种超媒体和超文本存储信息的系统称为超媒体系统或超文本系统。超媒体及超文本信息检索就是基于超媒体系统和超文本系统而进行的信息检索,包括基于浏览和基于提问两种方式。

(三) 文献信息检索的原理

犹如商品从生产到进入市场的流通,乃至最终的消费要经历一系列环节和过程一样,作为特殊智力“商品”的知识、信息、文献,从其产生到被吸收和利用,同样有一个如图 1-4 所示的过程。



图 1-4 检索原理示意图

从图 1-4 不难看出文献信息检索是两个既相对独立又密切联系的互逆过程。

对相关文献信息进行有选择性收集,对其内容进行分析提炼,借助于词表等工具赋予其若干不同形式的检索标识(如主题词、分类号等),将众多的经过加工后的文献信息按照其检索标识的特征(如分类或词序)组织为一个有序化的文献信息的集合并存储于磁盘等特定载体之上,便形成了所谓的文献信息检索系统。这个过程被称为文献信息的存储。

根据特定的信息需求,对其进行分析以形成一个明确的检索课题,即确定涉及几个概念、这些概念的内涵和外延如何、概念之间的联系或关系是什么、用什么样的标识表达这些概念等,根据需求特点选择适当的检索系统(如 PubMed 或 Web of Science)并选择相应的检索途径(如关键词、主题词、分类),在此基础上借助于检索技术(如处理概念之间逻辑关系的布尔算符、词间位置关系的位置算符等)将检索内容组合成能够被检索系统识别和运行的检索表达式,以此输入检索系统以便从中查找所需文献信息,这个过程被称为文献信息的提取。

实际上,文献信息的提取或查找所遵循的规则恰为最初存储时的诸多做法,即以什么方式存进去,也只能以同样的方式取出来。所以,广义的检索是包括系统的建立之存储和系统的利用之查找这两个互逆过程。而狭义的检索只涉及后者。作为医学专业人员,着眼于如何快捷而准确地获取和利用文献信息以做好科研、临床或教学工作,可将学习范围限于后者。但从上述检索原理的阐述中不难看出,如果对检索语言、词表、标引规则等文献信息的存储缺乏必要的

了解,实难做到检索的高效、快捷,更谈不上以查全率与查准率为标志的检索质量。

(四) 文献信息检索的意义与作用

作为一种以确切信息需求为导引、以信息检索系统为基础、以多种检索方法为手段、以信息思维为核心、以信息的获得为目的智力活动,文献信息检索有着多重的蕴意。

1. 有助于实现知识更新 随着科技的发展及其步伐的不断加快,知识陈旧的周期在不断缩短,新知识的涌现亦在不断激增。只有不断学习新知识以实现知识更新,才能顺应时代的发展。医学界早已倡导的毕业后教育、继续医学教育和终身学习的理念,绝不是抽象的观念,而是淋漓尽致地体现在医学工作者的专业发展和医学实践之中。作为自主获取知识信息的最重要途径,文献信息检索会有助于专业人员不断地汲取新知识、新方法和新技术。

2. 有助于解决实际问题 科学的魅力在于揭示自然和人类社会的奥秘,知识的意义在于解决人类生存与发展中遇到的新问题。作为呵护生命、增进健康的医学实践,每天都面临着新的情况、新的疑惑、新的挑战。应对的唯一策略就是及时、有效地运用已有的知识,因为“知识就是力量”。近20年来,在临床医疗实践中倡导的循证医学的理念、形成的循证实践的模式,所强调的“证据”就是新知识的结晶,“证据的获取”就是文献信息检索的体现。为此,PubMed等检索系统,为循证医学实践,提供了专门的途径和方法。善于获取与利用文献信息以解决实际问题,是医学工作者实践能力的重要组成部分。

3. 有助于推进科研进程 医学,承担着双重使命——运用医学知识服务于社会大众和创造医学知识以推进医学科学发展。科学研究,对于医学的发展乃至医疗服务水平的提升至关重要。科学研究强调以继承性为基础(在前人和他人成果上的进一步发展)、以创新性为灵魂(避免重复他人工作)、以科学性为保障(强调选题、设计与论证以科学事实和科学理论为根据),而这一切无疑以充分掌握、有效利用现有相关知识信息为前提。在文献信息的字里行间,会发现新的科学事实、新的学术观点、新的方法路径、新的技术手段。有避免重蹈覆辙的前车之鉴、有事半功倍的通幽曲径,更有柳暗花明的灵动启迪。因此,统领驾驭文献信息的能力,是科学能力的一种集中体现。检索方法与技术,乃至内化形成的检索能力,恰为荡漾于知识海洋中的一叶扁舟。

4. 有助于开发信息资源 资源的意义在于可被利用,资源的潜在价值更有赖于被利用的程度。作为支撑现代社会发展三大重要资源之一的信息资源,对其开发利用的效率与水平决定了发展的状态。医学的进步、医疗卫生事业的发展,同样以开发利用医学信息资源为前提。图书馆里的医学书刊、数据库里的医学文献、网上流动着的医学信息,有赖于医学专业人员的开发。此中的开发,超越了一般意义上的利用,强调的是开拓利用的范围、发掘利用的层次,使蕴含其中的价值得以最大限度地发挥效能。目前,学术界热衷的数据挖掘,就是旨在透过知识信息显现出来的当前价值,深入到其隐含着的潜在意义。如是,文献信息检索,亦就成为信息开发最重要的途径。

通过上面的阐述,不难看出以往所述的信息素养,不是字面意义上的名称、不是抽象层面的概念,而是鲜活而具体地体现于上述信息活动之中。因此,文献信息检索,是培养医学生的信息能力,提升其信息素养的最有效方式。

(于双成)

第二节 文献信息检索系统

人类生存与发展需要文献信息,而文献信息常常呈无序、散在的混乱状态,这种无序状态的文献信息不能成为有效的资源。于是,为满足人类信息需求而设计生成的人工系统——文献信息检索系统应运而生。所谓文献信息检索系统,是指根据人类社会对知识信息的普遍性需

要,由一定的硬件设备和软件条件所构成的,具有选择、整理、加工、存储和检索文献信息功能的有序化信息资源集合体。

一、文献信息检索系统的构成

系统是相互联系、相互作用的诸多元素的综合体。文献信息检索系统作为一种具有选择、整理、加工、存储、检索文献信息的开放式的多功能系统,自然有其构成要素、结构关系和整体功能。从信息检索系统的构成性要素来看,是由以下几个基本要素构成的。

1. 文献信息资源 是系统存储与检索的对象,是经过加工、整序后的序列化信息集合,也常称为数据库。它是系统提供信息检索服务最重要的物质基础,其内容可以是全文信息,也可以是其浓缩替代品的二次信息,还可以是事实或数值等各类信息,信息表述的形式可以是文字、图像、声音等多种形式。

2. 硬件设备 是实现信息存储、管理及检索任务的各种物理设施的总称,它构成系统生成与运行的物理结构基础,集中体现着信息依存性的特征。主要包括以大容量存储和高速度运算为标志、以高性能计算机为核心的主计算机或服务器;外部存储器、输入输出设备等外围设备;借助于通讯设施在内的数据处理或数据传送的其他设备。

3. 软件条件 系统内部的复杂运行和系统外部的诸多联系,均有赖于人为赋予并为计算机所识别和执行的各種程序,它构成了系统的灵魂与核心。主要包括系统软件和应用软件两大类。系统软件支持系统的基本运行,是各种不同类型应用软件得以实现其功能的基础,如操作系统、编译程序与汇编程序;应用软件是直接使用的,能够完成某种特定功能的程序。

由于不同信息检索系统的服务范围、宗旨和目的不尽相同,因而在系统的设计思想与总体结构上可能有所差异,其系统软件和应用软件的具体结构也就自然不同。正因为后者的差异,才使各系统呈现出别样的风格、具有了独特的功能,以适应和满足人们多样性的个性化需求。

4. 人力资源 是系统生成与运行中的能动因素。就目前来讲,计算机只能按照人的意志被动执行人类赋予的各种任务指令。尽管可有一定程度的创造性,但从根本上讲,它只能模拟人的思考,而不能替代人的思维。因此,无论是系统的分析与设计,还是数据的采集与整理,以及系统的维护与更新等,都有赖于相关专业技术人员的创造性工作。人力资源作为系统的一个必要组成部分,是系统其他各要素得以组合为一定结构、按一定逻辑规则运行、呈现一定整体性功能的关键之所在。

上述四方面内容构成了文献信息检索系统的空间结构,是系统运行的最基本前提。

二、文献信息数据库的结构

文献信息数据库,是将作为文献信息检索系统之最基本物质基础的信息资源,经过加工整序后形成一定的结构体系,特别类似于一个庞大的存放有序的信息“仓库”,由此而得“数据库”之名。国际标准化组织《文献与信息术语标准》(ISO/DIS5127)为数据库的定义为:数据库是指至少由一个文档(file)构成,并能满足某一特定目的或某一特定数据处理系统需要的一种数据集合。美国著名信息检索专家 Williams ME 为数据库所下定义是:数据库是指包含书目及与文献有关数据的机读记录(record)的有序集合。可以从如下两个维度来讨论数据库的结构。

(一)从数据库组成的层次来划分

存储有数以万计文献信息的数据库,是按如下几个层次构建的:一个数据库划分为若干个文档;一个文档存储一定数量的记录;每一个记录由若干个字段所组成。

1. 文档 数据库可按所属学科专业的不同或依年代时间范围,划分为若干个文档。例如,世界上最大的国际联机检索系统——Dialog 系统,拥有包括自然科学、人文社会科学及经济、商贸信息在内的 450 多个数据库,划分为 630 多个文档,我们常用的 MEDLINE 属于其中,依年代

进一步划分出的 154(1990 年以来)和 155(自 1950 年以来)两个文档。从这个意义上讲,可以将文档理解为组成大型数据库的子库,是数据库多层次结构中的一级组成部分。

2. 记录 每个文档是由许多的记录所组成。每一条记录都代表着经过加工处理的一篇文章或一则信息,它揭示了文献的内容和形式特征(如文献的题录、文摘、著者,以及标引的关键词或主题词等)。因此,记录是构成数据库的最基本数据单元。

3. 字段 每个记录一般由若干个描述性字段所组成。每个字段描述文献信息的某一内容或形式特征,即数据项,并且有唯一的供计算机识别的字段标识符(field tag),如篇名字段(TI)、著者字段(AU)、文摘字段(AB)、来源字段(SO)、主题词字段(MH)等。

通常情况下,检索是从某一个角度(如著者、主题、篇名)着手的,即从相应的著者字段、主题词字段、篇名字段入手查找相关内容。每一个字段不仅是构成记录的基本单元,在数据库检索时,也是一个检索点或检索入口(searching entry)。凡是可以作为检索点的字段称为可检字段,是检索得以实现的基础。

数据库的结构,可用英文单词的单复数形式更直观地表述为图 1-5:



图 1-5 数据库的结构

(二) 从数据库组成的方式来划分

文档的另一层含义是指信息在数据库内的组织形式。不同的组织方式,形成了具有不同功能作用的文档形式。

1. 顺排文档 文档中全部记录按一定顺序依次存储于磁带或磁盘等介质所构成的存储器中,且有一个唯一性的物理位置标识即顺序号。该文档中存储的是替代一则信息、一篇文献的完整记录,它构成数据库的主体,所以又称为主文档。由于存取的记录均依赖于有章可循的顺序,便形成了记录之间单一线性的关系,所以又称为线性文档。

2. 倒排文档 考虑到用户检索信息时往往是从信息或文献的某一特征入手,如著者、主题、出处等。因此,数据库建设时,将这些具有检索意义的特征一一抽出并附带其在主文档中的位置标识,按照一定顺序重新排列形成一种新的文档。这些特征在新文档中的排列顺序可能有别于主文档中记录存放的顺序,故而称之为倒排文档。一个数据库只有一个主文档,但可根据需要编制若干个倒排文档,如著者文档、主题词文档、篇名文档等。

检索时,系统按输入检索提问的内容与形式首先到相应的各种倒排文档中搜索,找到相匹配的检索标识后,依据其附带的存放于主文档中的位置标识再到主文档中调取相应的记录。这种在有限范围内有针对性的搜索,犹如按图索骥,有别于在整个主文档中的海里捞针,极大地提高了检索的效率。

犹如一本书的索引是为查找书中正文内容服务、二次文献是为查找一次文献服务一样,倒排文档是为查找主文档中文献信息内容服务的,它们是从主文档中派生出的附属结构。主文档是以一个个记录为存储和检索的基本单位,而倒排文档中是以记录中的内容或形式,这些体现文献信息不同属性且具有检索意义的特征标识为排列依据。因此,检索系统的数据库才具有了前面所讲的结构层次性。主文档和倒排文档的关系可用图 1-6 表示,其中实线箭头表示倒排文档的生成,虚线箭头表示检索过程。



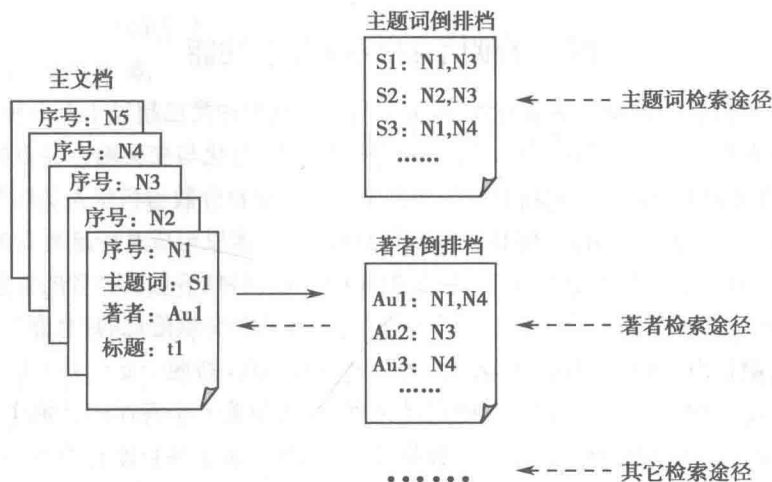


图 1-6 主文档和倒排文档的关系示意图

三、文献信息检索系统的分类

文献信息检索系统的核心是数据库，而数据库本身又常以收录内容及功能的不同划分为如下几种类型。

1. 书目型数据库 (bibliographic database) 主要是指二次文献数据库，包括各种机读版的题录型、文摘型的数据库。如《中国生物医学文献数据库》(SinoMed)、MEDLINE 等。它们提供了可满足用户多种信息检索需求的有关文献的各种特征，如文献的篇名、著者、出处(包括刊名、年、卷、期、页码)、摘要、收藏单位等。不仅可以告知用户其所需文献的线索——题录(包括篇名、著者、出处)，更可以提供整篇文献内容浓缩的替代品——文摘。因此，这是文献检索中最常用的一种数据库。

2. 事实数值型数据库 (fact-date database) 主要为用户提供有关事物、人物、机构等方面的事实性信息和数值型数据。例如万方数据资源系统中的事实型数据库《中国科研机构数据库》《中国科技信息机构数据库》《中国高等院校及中等专业学校数据库》，美国国立癌症研究所的关于癌症治疗信息的医生咨询数据库(PDQ)，反映药物处方信息的《医师案头参考书》(Physicians' Desk Reference, PDR)等。电子化的参考工具书，如词典、百科全书、指南等也属于事实型数据库的范畴。

3. 全文型数据库 (full-text database) 是将文献全文的全部内容转化为计算机可以识别、处理的信息单元而形成的数据集合。全文型数据库对文献的字、词、句，乃至段落等进行更深层次的编辑加工，允许用户采用自然词语以及截词、邻近算符等匹配方法，方便快速地查到所需的文献，并能直接获取文献原文。例如 CNKI 的学术文献总库、万方数据资源系统中的数字化期刊、Elsevier Science 电子期刊全文库。因此，全文型数据库是集文献线索的查询和文献全文的提供于一体，实现了毕其功于一役的一站式信息服务。

4. 知识型数据库 (knowledge database) 一种疾病，不仅有一个概念性的界定，而且涉及病因、病理、诊断、治疗等相关知识；一种药物，不仅有其化学结构上的名称或命名，而且有药理作用、药效动力、副作用、适应证、使用方法等相关知识；一项检查，不仅有其概要性描述，而且涉及原理、检测值的解释、临床意义、操作方法、正常值范围、适用范围等一系列相关知识。以这些疾病、症状、药物、检查项目等概念为知识点或检索点，将这些概念的知识解析内容，以概念的相关语义关系为纽带编制成一个具有知识结构特征和层级范围的立体网状的知识库。与以往的文献信息系统不同，检索到的不是文献，而是对概念的知识解读。例如中国疾病知识总库 (China Disease Knowledge Total Database) 所包括的疾病数据库、药品数据库、辅助检查数据库等多种知识库。

四、文献信息检索系统的功能

1. 汇集文献信息 以期刊文献为例,世界上科技期刊的种类已超过十万余种,此外还有众多的其他种类的正式和非正式出版物,加之现代科技专深、分化与交叉融合并重的趋势日渐强劲,都使得科技文献的发表及相关信息的发布异常分散。这种分散与科研人员以学科为基础的信息需求之间形成了矛盾。为此,将相关文献信息收集、汇聚以形成某一学科专业的文献信息数据库,或形成适应若干学科专业的综合性数据库,便成为科技信息工作者的重要使命。因而也就赋予了文献信息系统第一个功能——作为汇集、存储相关文献信息的“仓库”。

2. 组织文献信息 仍以期刊文献为例,每一种科技期刊,按照一定的办刊宗旨、报道范围和出版规则,选录、编辑、刊发一定数量的科技文献,使之成为一个有序的文献体系。然而,将数千种期刊中数以万计的文献汇集于一个数据库中,无疑打破了各自原有的秩序,成为一种无序的混乱。而没有经过组织的信息算不上是一种资源,混乱无序的文献信息无法得以有效开发和利用。因此,需按照一定的规则(如按照文献论及的主题内容或是按照其所属的学科体系),将这些文献信息重新加工组织成一种井然有序的文献信息集合,这是文献信息系统的核心即数据库建设的关键环节。由此可见,文献信息检索系统的第二个功能就是对文献信息的整序,亦称之为文献信息的组织。

3. 检索文献信息 通过前面的阐述,不难看出,数据库乃至整个文献信息系统的建立,旨在解决文献信息的散乱无序性与用户的特定需求之间的矛盾。对文献信息的整序组织,是以赋予每一篇文献、每一则信息以多维检索标识为标志。这些检索标识,形成了在文献信息的迷宫中寻奇探宝的通幽曲径。例如,本书后面章节将要介绍的许多数据库,均可从著者、篇名、主题、关键词等诸多角度查找相关文献信息。这便是文献信息检索系统的第三个,也是最具特征性的功能——可查检性。

上述三项功能,存在着这样一种内在的逻辑关系:收集文献信息是基础、组织文献信息是手段、检索查询文献信息是目的。

(于双成)

第三节 文献信息检索技术

文献信息检索技术亦有广义和狭义之分。广义的文献信息检索技术包括信息组织、数据库建设在内的涉及信息科学、情报学、计算机科学等诸多学科领域的技术方法,而狭义仅指从现有的信息资源中提取相关信息的技术方法。

文献信息检索技术的发展经过四个阶段。一是手工检索阶段:检索方式主要以手工操作为主,这种检索不仅费时、费力,而且检索效率也很低。其中包括纸质文献的检索和缩微式检索。二是脱机检索阶段:信息检索逐步实现了计算机检索中的单机批处理检索。包括计算机可读文献磁带和磁盘检索以及光盘数据库检索。机读磁带、磁盘检索开始实现了一种输入多种输出。光盘数据库比磁带和磁盘有更大的存储空间,且存储速度更快。这是计算机检索的第一阶段。三是计算机联机检索阶段:进入20世纪70年代,计算机软、硬件技术不断进步,分组数字通信技术和实时操作技术发展迅速,出现了一台主机带多个终端的系统,用户可以利用计算机检索终端设备,通过拨号、电信专线及计算机互联网络,从联机服务中心的数据库中检索出自己所需要的信息,从而实现了计算机联机检索。它属于计算机检索中的第二阶段。四是网络化信息检索阶段:网络信息检索是指利用计算机设备和国际互联网检索网上各服务器站点的信息。目前国内网上数据库有:国家科技图书文献中心、中国学术期刊数据库、万方数据库、重庆维普数据库、中国专利文献数据库、中国标准数据库以及超星数字图书馆、书生之家等全文数据库;国

外网上数据库则更多,如美国科学索引(SCI)、工程索引(EI)、化学文摘(CA)、Clinicalkey 数据库、SpringerLink 数据库等。

一、文本信息检索技术

文本信息检索是一种较为简单的准确匹配模式,主要的文本信息检索技术包括以下几种。

1. 布尔逻辑检索技术 是计算机文献检索中最常用的检索技术。在文献数据库检索系统中,用户的信息需求是通过检索提问式表达的。布尔逻辑算符在检索提问式中起着逻辑组配的作用,它能把一些具有简单概念的检索单元组配成一个具有复杂概念的检索式,用以表达用户的信息需求。由于系统中采用的逻辑组配算符是布尔代数中的逻辑符 AND、OR 和 NOT,故称布尔逻辑算符。这三种逻辑算符各表示不同的逻辑思想,依此将用户提问转换成逻辑表达式。计算机可根据表达式所限定的各运算项(或称检索词,包括主题词、关键词、分类号、著者、化学物质登记号、公司代码等)之间的关系,确定命中文献的基本条件与查找途径,以便同文档中各篇文献的标识进行匹配。凡符合条件的文献即为命中文献,予以输出。

布尔逻辑具体检索时,是通过三个布尔运算符来实现其功能的: and、or、not, 分别表示逻辑与、逻辑或、逻辑非三种逻辑运算关系,如图 1-7 所示。

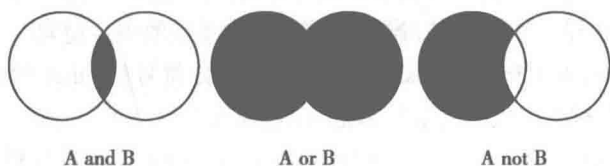


图 1-7 布尔逻辑运算

(1) 逻辑“与”,“and”: 检索表达式为“A and B”。数据库中同时含有检索词 A 和 B 的文献为命中文献。其作用是缩小检索范围,提高查准率。例如查找“胰岛素治疗糖尿病”的文献,检索式为 insulin(胰岛素)and diabetes(糖尿病)。

(2) 逻辑“或”,“or”: 检索表达式为“A or B”。数据库中的文献中含有检索词 A 或 B 的文献为命中文献,其作用是扩大检索范围。例如查找“肿瘤”的文献,检索式为 cancer(癌)or tumor(瘤)or carcinoma(癌)or neoplasm(新生物)。

(3) 逻辑“非”,“not”: 检索表达式为“A not B”。数据库中凡含有检索词 A 而不含 B 的文献为命中文献,其作用是缩小检索范围,提高查准率。例如查找“动物的乙肝病毒(不要人的)”的文献,检索式为 hepatitis B virus(乙肝病毒)not human(人类)。

当一个检索表达式含有多个布尔算符时,执行的顺序为 not 优先运算, and 其次, or 最后。如要改变,可用圆括号改变运算顺序,将需要优先运算者置于圆括号中。例如查找“维生素 C 或维生素 E 对糖尿病患者肾脏的保护作用”,检索式为:(维生素 C or 维生素 E)and 糖尿病 and 肾。

2. 截词检索(truncation search) 是利用截断的词的一个局部,即截词进行检索的技术。截词检索能自动地对同一概念检索词的不同词尾变化、词根相同等一类提问词进行检索,并自动用“or”连接截词的检出结果,避免了文献的漏检和逐词键入的麻烦,有效提高查全率。

(1) 按是否采用截词符号,截词检索分为 2 种:

1) 采用截词符号: 常用的截词符号有“*”、“?”、“\$”等。

2) 不用截词符号: 系统支持自动截词,如在 PubMed 的基本检索方式下输入“Smith J [au]”,系统能自动检索出包含 Smith J 的著者(如 Smith JB、Smith JS 等)撰写的文献,此外,有些检索系统也支持检索包含同一词根的一组词,如在 EI 的快速检索中输入“management”,系统能将 managing、managed、manager、manage 等检索出来,这种检索技术通常称为词干检索(stem search)。

(2) 按截断位置的不同,截词分为3种:

1) 右截词:截词符在检索词末端,用于检索词头相同的一组词,如 cell*。

2) 左截词:截词符在检索词的最前端,用于检索词尾相同的一组词,如 *mycin。

3) 中间截词:也称通配检索,截词符(通配符)出现在检索词的中间,代表若干个字母,如 m*n。

(3) 按截断的字符数量的不同,截词检索分为2种:

1) 无限截词检索:即不限制被截断的字符数量的检索。“*”常用于无限截词(*=0~n 个字符)。例如输入“pain*”,可同时检出 pain、painful、painless、painkiller 等。

2) 有限截词检索:即限制被截断的字符数,用以替代一个字符或不替代任何字符的检索。“?”常用于有限截词(?=0~1 个字符)。例如输入 wom?n,可同时检索出 woman 和 women 等。

3. 邻近检索(proximity searching) 是用来表示检索词与检索词之间位置关系的检索技术,通常用位置算符来实现。采用的位置算符有 near, with。

(1) “near”:表示检索词位置相邻,检索词出现的顺序可前可后,如 gene near apoptosis。

(2) “with”:表示检索词位置相邻,且两词出现的顺序与输入顺序一致,如 liver with cancer。

4. 限定检索(limit searching) 又称限定字段检索(limit field searching),是利用检索词出现的字段进行的检索。几乎所有计算机检索系统均支持限定检索,用户可以指定检索某一字段或某几个字段使得检索结果更为准确,减少误检。检索系统的限定检索会采用缩写形式的字段标识符(如 TI 表示 title, AD 表示 address 等)甚至采用限定符号(如 in),如 SinoMed 的高级检索方式下检索“篇名中含有高血压的文献”,需键入:高血压 in ti。

5. 扩展检索(expand searching) 是同时对多个相关检索词执行逻辑或检索的技术,即当用户输入一个检索词后,系统不仅能检出该检索词的文献,还能检出与该检索词同属于一个概念的同义词或下位词的文献,如 PubMed 中主题词的扩展检索。

6. 加权检索(weighing searching) 检索运算时不仅查找检索提问词,而且考虑并计算提问词的权值,当权值之和超过阈值的记录才能在数据库中被检出来;在某些数据库中表现为仅检索主要概念主题词,如 MEDLINE;而有些数据库表现为词频检索,如中国学术期刊全文数据库。

7. 精确检索和模糊检索 精确检索(exact searching)是指所检信息与输入的词组完全一致的匹配检索技术。在许多系统中用引号来表示,如检索“acute pancreatitis”,此时只有包含与 acute pancreatitis 完全相同的词串的文献才能检索出来;而模糊检索(fuzzy searching)允许所检信息与检索提问之间存在一定的差异,如检索 acute pancreatitis,可检索出 acute necrotizing pancreatitis、acute gallstone pancreatitis 等,即只要包含 acute 和 pancreatitis 两个词串的文献均能检索出来,并不要求 acute pancreatitis 一定按输入顺序相邻。

8. 跨库检索(cross database searching、one-stop searching) 是一次对多个数据库同时进行检索的技术。跨库检索能为用户提供统一的检索接口,将用户检索需求转化为不同数据库的检索表达式,并发地检索本地和广域网上的多个分布式异构数据库,并对检索结果加以整合,以统一的格式将结果呈现给用户,在此过程中用户不必担忧这些数据库的协议、平台、产品或生产商,因而能够减轻用户学习检索不同数据库的负担,有效节省检索时间。

9. 相关信息反馈检索 是将与已检结果存在某种程度相关的信息检索出来的检索技术,多由检索系统自动进行检索,如 Google (<http://www.google.com>) 的“类似网页”、PubMed 的“Related Articles”。

10. 智能检索 即自动实现检索词、检索词对应主题词及该主题词所含下位词的同步检索。如中国生物医学文献服务系统(SinoMed)的智能检索。PubMed 的“自动词语匹配检索”也属于智能检索,也称自然语言接口技术。



二、多媒体信息检索技术

随着互联网技术与计算机图像数码技术的飞速发展,人们接触到大量的图像、视频、音频等信息。因此,怎样从网络数据库中快速查出相关多媒体信息数据,成为数字化图书馆等重大研究项目中的关键技术。

早年对于多媒体信息的处理是转化为基于文本描述的检索,通过对图像等信息进行手工注解,然后利用文本检索技术进行关键字检索。这种检索方式存在两种缺陷。一是手工注解图像库的工作量太大;二是由于文本注释不能完全表达图像所包含的丰富内涵,造成了文本注解不可避免地存在主观性和不精确性。因此,需要开发出一种新型的技术来检索多媒体数据,基于内容的多媒体信息检索技术应运而生。

基于内容的检索(content based retrieval, CBR),主要是对多媒体对象的语义、媒体的视觉特征或听觉特征进行检索。它利用图像处理、模式识别、计算机视觉、图像理解等学科中的医学方法作为部分基础技术,直接对图像、视频、音频内容进行分析、抽取特征和语义,建立内容特征索引以供检索。

基于内容的图像检索(content based image retrieval, CBIR)使用了可以直接从图像中获得的客观的视觉内容特征,如颜色、纹理、形状等来判断图像之间的相似性。这种方法成了现有图像检索技术研究的主流。其主要研究内容是在数字图像处理基础上的视觉特征提取、多维索引以及检索系统设计。这类系统主要支持基于范例检索(example-based retrieval)、基于草图检索(sketch-based retrieval)和随机浏览及其组合的工作方式。

图像的内容特征可以分为两类:低层视觉特征和高层语义特征。低层视觉内容主要包括颜色、纹理、形状、空间关系等特征,可以通过特征提取获得。以图像的低层视觉内容为基础,基于多媒体内容的图像检索的关键技术如下:

1. 基于颜色特征的检索 颜色内容包含两个概念,一个对应于全局颜色分布,一个对应于局部颜色信息。按照全局颜色分布来索引图像可以通过计算每种颜色的像素个数并构造颜色灰度直方图来实现,这对检索具有相似总体颜色内容的图像是一个很好的途径。局部颜色信息是指局部相似的颜色区域,考虑了颜色的分类与一些几何特征。

2. 基于纹理特征的检索 纹理可以视为某些近似形状的近似重复分布,纹理描述的难点在于其与物体形状之间存在密切的关系,千变万化的物体形状与嵌套式的分布使纹理的分类变得很困难。一种方法是利用纹理特征的共生矩阵表,即根据像素间的方向和距离构造一个共生矩阵,然后从共生矩阵中抽取有意义的统计量作为纹理表示。另一种是从视觉的心理学角度提出了“纹理表示”方法,表示的所有纹理性质都具有直观的视觉意义,这使得“纹理表示”在图像检索中极具吸引力,且可提供一个更友好的用户界面。

3. 基于形状特征的检索 一般说来,形状的表达可分为基于边界的和基于区域的两类,前者使用形状的外部边界,而后者使用整个区域。

4. 基于组合的多特征检索 结合图像的颜色、纹理、形状、空间关系等多种图像特征,全面描述图像内容的检索方法。该方法可以达到不同特征的优势互补的效果,提高检索的灵活性和检索系统的性能。

5. 基于语义特征的检索 基于图像语义进行检索的目的是希望能够更好地从用户的角度出发,查找出与用户理解相一致的图像。用户对图像内容的语义理解必须通过一定的方式表达出来,以提供给应用系统处理。语义内容的表示方法有多种,如语义网络、数理逻辑、框架方法等。为了克服基于简单视觉特征的图像检索方法的不足,人们提出了基于语义的图像检索方法,与基于低层物理特征查询不同,语义特征查询是基于文字的查询,包含了自然语言处理和传统图像检索技术,这种检索方法的目标是最大限度地减小图像简单视觉特征与丰富的语义之

间的语义鸿沟,缩小语义鸿沟的办法有2种,由高层语义导出低层特征和由低层特征向高层语义的转换。图像语义具有模糊性、复杂性、抽象性,一般包括3个语义层次:特征语义、目标和空间关系语义、高层语义。特征语义就是图像的颜色、形状、纹理等低级视觉特征;与视觉感知直接相连,目标语义和空间关系语义需要识别和提取图像中的目标类别目标之间的空间位置等关系,涉及模式识别和逻辑推理的相关技术;高层语义主要涉及图像的场景语义(如海滨街道室内等)、行为语义(如表演超越进攻等)和情感语义(如平静和谐振奋等)。一般而言,高层的图像语义往往建立在较低层次的语义获得的基础上,并且,层次越高,语义越复杂,涉及的领域知识越多。基于语义的图像检索一般指的是基于目标和高级语义的图像检索方法,在原有检索系统中加入高级语义到低层特征的转化,可以在不改变现有的图像特征库和匹配方式的情况下实现基于语义的图像检索,其基本框图见图1-8。

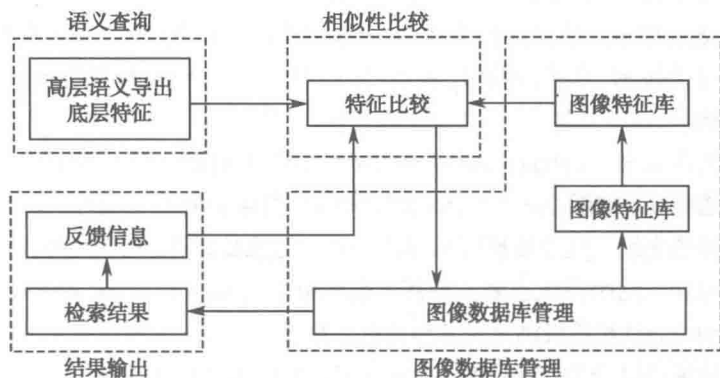


图 1-8 基于语义的图像检索系统框图

6. 相关反馈技术 颜色特征、纹理特征、形状特征均属于图像的低层特征,图像低层特征和高层语义间存在着很大的差距,仅基于图像低层特征很难检索出令用户满意的结果。为了解决这个问题,一方面需要研究出更好更有效的图像表示方法,另一方面可以通过人机交互的方式来捕捉和建立低层特征和高层语义之间的关联,这就是所谓的相关反馈技术。基于内容的检索方式的相关反馈技术大致可分为四种类型:参数调整方法、聚类分析方法、概率学习方法和神经网络方法。

基于内容的多媒体检索技术与传统数据库技术相结合,可以方便地实现海量多媒体数据的存储和管理;与传统 Web 搜索引擎技术相结合,可以用来检索 html 网页中丰富的多媒体信息。

三、可视化信息检索技术

随着数字化信息的指数增长,存储的数据量越来越大,用户在进行信息检索时系统往往会返回一个大型的数据集,过量的信息已经逐渐成为信息有用性的一道障碍。显然需要一种新技术来帮助用户从海量文献信息中查找和收集有用的资料,可视化就是这样一种将检索过程简单化、高效化的技术工具。

信息可视化(information visualization)是将抽象数据用可视的形式表示出来,包括访问的结果以及数据各部分之间的关系,用于指导和加速查找的过程。信息可视化主要有层次信息结构可视化、多维数据结构可视化、网络结构、运行状态、浏览历史及网络用户的可视化等。

可视化检索(visual retrieval)是信息可视化的一种。是指把文献信息、用户提问、各类情报检索模型以及利用检索模型进行信息检索的过程中不可见的内部语义关系转换成图形,在一个二维或三维的可视化空间中显示出来。换句话说,将数据库中的文献及它们之间的关系看作一个抽象的信息空间,该空间中包含成千上万篇文献,且文献间相互关联,甚至文献的标引词之



间也存在某种联系。但由于数据库的高维性,使得这些关系是不可见的,可视化将这些联系用可见的方式表示出来。

相对传统检索而言,可视化检索作为一种非传统信息检索机制,有其自身的优势。

1. 检索过程透明化 对用户来说,传统信息检索的检索过程始终是不可见的“黑盒子”(back box),用户无法对具体处理操作进行控制,也不知道他们输入的检索词在系统内部是如何工作并输出结果的。可视化检索的过程则是透明的,通过对文献与检索要求间内部关系直观的图形显示,帮助用户选择相关文献,并增加了用户对检索内部过程进行控制的新功能。

2. 检索结果输出高效性 传统信息检索提供文本形式的线索,占用大量屏幕空间,大大增加了用户查找所需信息的难度。可视化检索中用点或符号来表示数据库中的文献,几乎所有检索结果都能在一个屏幕中同时显示出来,提供整体浏览。用户可任意选择其中的某点,了解对应文献的详细资料。

3. 更为有效的检索结果排序机制 虽然传统检索系统都会按其各自的相似匹配算法对文献与检索问题是否匹配进行判断,并计算出两篇文献之间的相似度,但用户很难由此判断出文献与多个检索词的检索提问中哪个检索词相关度较高,也很难将不同检索式所返回的检索结果进行比较。可视化检索用 2D 或 3D 的图形显示代替 1D 的线性输出,对图形中可视化的文献分布情况进行分析,进而确定一个词对于标引和检索的价值,多角度多层次反映文献与检索要求、文献与文献之间的语义关系。

4. 有效的用户反馈机制 增强了用户与系统间的交互作用,允许用户对检索结果进行动态调整和过滤,甚至根据自己的检索需求标记出相关文献与非相关文献,改变了传统检索过程中不考虑用户检索偏好,缺少根据用户具体要求的变化进行动态检索的状况,进而提高了查准率。

信息可视化检索可以分为两种:

1. 文档的可视化 文档信息是人类记忆的延伸,为了理解文档及其中的信息,需要阅读大量文本。文档之间的主题或内容的相关性对使用者来说是非常重要的,例如在因特网搜索信息时,需要快速把握成千上万的搜索结果和搜索条件之间的相关性,从而迅速找到真正需要的信息。美国西北太平洋实验室的科学家提出 SPIRE (spatial paradigm for information retrieval and exploration), SPIRE 可以用于任何类型的文档集合,确定文档之间的关系,使之可视化,对于信息检索和知识挖掘非常有用。

2. Web 的可视化 互联网上 Web 信息不计其数,这些信息分布在世界各地的网站上,网站通过文档之间的超级链接连接在一起,如何方便利用 Web 信息,成为亟待解决的问题。信息可视化使得用户与互联网信息之间的交互变得越来越容易。

自可视化技术应用于信息检索以来,专家们已经研究开发出多种可视化信息检索系统,未来可视化检索的工作重点应为新系统的继续开发和原有系统的不断改进,另一方面则应参照现有检索系统的评价方案,逐步建立起可视化检索系统的评测体系,以促进信息检索可视化的进一步发展。

(罗爱静)

第四节 文献信息检索途径与策略

一、文献信息检索途径

信息检索系统对于所收录信息的加工、整序是以其内容和形式特征为依据。以印刷型文献信息为例,涉及文献信息内部特征的有文献论及的主体内容、所属的学科范围;涉及文献外部特征的有著者姓名、载体类型、出版时间等。正是基于这诸多不同的特征,形成了多种相应的

倒排文档(或索引),而每一个倒排文档都构成信息检索系统的一个检索途径。

检索途径为检索系统的检索入口,即检索系统所提供的、用以查询获取资源的各种标识,在计算机检索系统中通常表现为字段检索。常见的计算机检索途径有以下几种。

1. 主题词检索途径 利用主题词(subject headings)作为检索入口来查找文献。主题词是经过优选和规范化处理的词汇,由主题词表来控制,常用的有美国国家医学图书馆(NLM)编制的《医学主题词表》。

目前支持主题词途径的检索系统主要有 Sinomed 和 PubMed,其中 Sinomed 提供的主题词途径是采用《医学主题词表》《中医药学主题词表》中的主题词或主题词/副主题词作为检索入口。

2. 自由词检索途径 通过从文献篇名、正文或文摘中抽取出来的能表达文献主要内容的单词或词组查找文献的检索途径。自由词用词灵活、自由、直观、符合用户习惯,因而自由词途径已成为目前检索系统中应用最为广泛的一种检索途径。但自由词因未经规范化处理,检索时必须同时考虑到与检索词相关的同义词、近义词等,否则容易造成漏检。如检索“艾滋病”时需要考虑其学名“获得性免疫缺陷综合征、AIDS等”。

3. 分类检索途径 是利用文献在分类体系中的位置(类目名称或分类号)作为检索入口查找文献的途径,可满足用户从学科、专业等内容出发获取文献的需要。如 Sinomed 提供了分类途径,以文献在《中国图书馆分类法》中的分类号作为检索入口来查找文献。

4. 著者检索途径 是利用文献上署名的作者、编者或机关团体名称作为检索入口查找文献的途径。在外文检索工具中,通常采取姓在前、名在后用首字母缩写的方式进行检索,如利用 WOS 著者途径检索夏家辉院士发表的文献时,需从“XiaJ”或“XiaJH”进行检索。并且文献检索出来后,还要根据著者的专业及其他特征做进一步鉴别,避免误检。

5. 题名检索途径 是利用题名(如书名、刊名、篇名)等作为检索入口查找文献的途径。

6. 序号检索途径 是利用文献特有序号作为检索入口查找文献的途径。文献中的序号主要包括 ISSN、ISBN、专利号、公开号、报告号、合同号、化学物质登记号等。

7. 机构检索途径 是以机构名称为检索入口查找文献的途径。不少数据库把机构名并入地址字段(address)。

8. 引文检索途径 是利用引文(citation)这一特征作为检索入口查找文献的途径。许多检索工具提供的引文检索途径允许以引文中出现的任何信息(如作者、题名、出处)等作为检索入口,如 Web of Science 的引文检索(cited reference search)。

9. 缺省检索途径 是指自动在检索系统预先设定的多个字段中同时进行检索。例如在 Sinomed 的基本检索界面直接输入检索词,系统默认在全可检字段中进行检索。

10. 其他途径 如《化学文摘》的分子式索引,《生物学文摘》的生物体索引。

二、文献信息检索策略

(一) 检索策略的含义

广义的检索策略是指用户根据检索需求选择相应的数据库、确定检索方式、检索途径及相应检索表达式进行检索的一系列操作或方案,是用户检索目标的体现。狭义的检索策略仅指用户确定检索表达式进行检索的系列操作。

检索表达式又称检索式或检索提问式,是用于表达检索需求的检索词或检索词的集合。主要包括以下几种情况:①单个检索词构成,如 Hypertension(检索高血压的相关文献);②单个检索词加运算符构成,如 Hypertension [Ti](检索篇名中含有高血压的相关文献);③多个检索词构成,如 Hypertension/therapy(检索治疗高血压的相关文献)。

用户制定的检索策略通常包含多个检索表达式,这些检索表达式由用户输入检索系统,系统检索之后经常会自动将检索情况记录在检索史(history)中,并按检索顺序给予每一个检索表



达式一个编号予以识别。用户通过检索史中可以浏览检索结果,还可利用编号调用、删除、保存和调整检索史中的检索表达式。

(二) 检索策略的构建

利用计算机检索工具进行检索时,用户构建检索策略的一般过程如下:

1. 分析检索课题,明确检索要求 分析检索课题的目的是使用户搞清楚其课题要解决的问题,即它所包含的概念和具体要求及它们之间的关系。这是制定检索策略的根本出发点,也是检索效率高低或成败的关键。

(1) 分析课题的主题内容:即分析课题的主题内容、所属学科性质,明确研究课题所需的信息内容,从而提出能准确反映课题核心内容的主题概念。

(2) 确定课题的文献类型:通过对课题进行主题分析后,确定所需信息的文献类型。如果属于基础理论性探讨,要侧重于查找期刊论文、会议论文。如果是尖端技术,应侧重于科技报告。如属于发明创造,技术革新,则应侧重于专利文献。如为产品定型设计,则需利用标准文献及产品样本。明确课题对检索深度的要求,弄清用户是需要提供题录、文摘还是原始文献。

(3) 确定检索时间范围:根据课题研究的起始年代和研究的高峰期确定检索的时间范围。

(4) 分析用户的检索评价要求:分析用户对检索评价指标是查新、查准还是查全。一般来说,若要了解某学科、理论、课题、工艺过程等最新进展和动态,则要检索最近的文献信息,强调查“新”;若要解决研究中某具体问题,找出技术方案,则要检索有针对性、能解决实际问题的文献信息,强调查“准”;若要撰写综述、述评或专著等,强调查“全”。

2. 选择检索系统,确定检索方法 根据检索课题的要求,选择最能满足检索要求的检索系统或工具书。一个计算机检索系统往往包含若干个数据库,进入系统后,常会有主体分类目录提供用户选择。一些内容相同的数据库也经常出现在不同的检索系统中。数据库的选择原则一般可概括成四个C,即“4C”原则:一是 content(内容),指数据库的内容、学科范畴、文献质量、数据库类型(如数值、事实、文摘、全文等)和文献来源(如期刊论文、会议论文、专利文献、科技报告等);二是 coverage(覆盖范围),指数据库的规模,涉及时间范围、地理范围、机构来源、文献量等;三是 currency(更新),指数据库更新的及时性,更新频率和周期;四是 cost(成本),即所需的检索费用,如数据库的使用费用、检索结果输出费等。

数据库选择好以后,在检索前,需阅读该库的说明,如出版机构的权威性、文献类型(期刊、会议、报告、专利等)、收录年限、服务功能等。

3. 确定检索途径,编写检索策略表达式 根据信息需求或检索课题的已知条件和检索要求以及所选定的信息检索系统所提供的检索功能,确定适宜的检索途径,如主题途径或关键词途径等。

检索途径确定后,编写检索策略表达式,即将选择确定的作为检索标识的作者姓名、主题词、关键词,以及各种符号如分类号等,用各种检索算符(如布尔算符、位置算符等)组合,形成既可为计算机识别又能体现检索要求的提问表达式。

4. 评价检索结果,修正检索策略 按照预定的检索策略进行检索,并对检索结果的相关性进行分析、评价,如果满足自己的检索需求,则根据需求采用一定的输出方式将检索结果输出。如对检索结果不甚满意,此时应对检索策略进行调整,以获取更好的检索结果。

(三) 检索策略的调整

检索策略的调整可概括为两种情形,即扩大检索(范围)或缩小检索(范围)。

1. 扩大检索 即指增加检出文献量,可采取以下措施:

(1) 重新选择数据库:选择多个数据库进行检索,或增加所检数据库的检索年限。

(2) 选择多种检索方式:不同检索方式有不同的特点,采用多种检索方式相结合,可以适当扩大检索。

(3) 重新选择检索途径: 如选择篇名字段检出结果较少时, 可选择文摘、组合字段或全文字段, 获取较多检出结果。

(4) 重新构建检索表达式: 如主题词检索时采用扩展检索, 选用所有副主题词或扩展下位副主题词; 自由词检索时考虑其同义词、近义词等, 并用 OR 相连; 采用截词检索, 且截词不宜太长; 去掉次要的主题词或非核心的检索词, 减少 AND 组配面; 对检索词不做限定; 采用邻近算符检索时, 不要过于严格; 浏览检索时选择多个相关类目; 采用模糊检索; 采用相关信息反馈检索等。

2. 缩小检索 是指减少检出文献量, 可采取以下措施:

(1) 重新选择数据库: 减少所检数据库的数量, 或减少所检数据库的检索年限。

(2) 选择最佳检索方式: 如选择高级检索或专家检索均可输入较多的限定条件, 可以适当缩小检索; 很多搜索引擎的分类目录是依据人工方式进行筛选分类的, 有时可缩小检索。

(3) 重新选择检索途径: 如全文字段检出文献较多时, 可重新选择在篇名、关键词和文摘等字段。

(4) 重新建构检索表达式: 如尽量采用主题词检索, 并借助主题词表选择更专指的下位词进行检索, 选择特定的限定词进行组配检索; 自由词检索时进行各种限定; 采用文献形式特征进行限定; 增加 AND 的组配面, 使检索表达式更为准确地表达检索需求; 减少 OR 的组配面; 用运算符(如 NOT、加权)排除带有干扰性的概念或不需要包含的概念; 词语检索时采用精确检索; 邻近算符检索时, 严格限制检索词之间的距离; 浏览检索时仅选择一个主要类目; 利用限定运算符(如 Filetype)限制检索某一特定类型或范围内的文献等。

综上所述, 整个检索的过程和步骤概括为如下流程(图 1-9)。

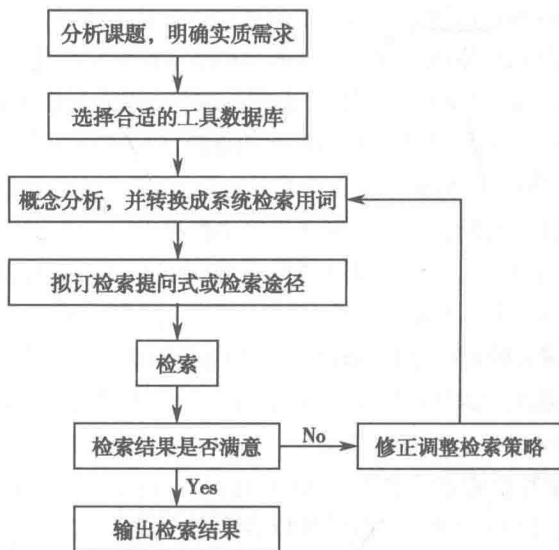


图 1-9 检索流程图

本章小结

本章介绍了文献信息的含义、类型、级别、特点及其社会功能, 阐述了文献信息检索的原理; 以及文献信息检索系统的构成和功能, 简要介绍了文献信息数据库的结构和类型; 对文本信息检索技术、多媒体信息检索技术和可视化信息检索技术以及文献信息检索途径和检索策略也进行了讨论。

(罗爱静)

思考题

1. 简述信息、知识、情报三者的联系与区别。
2. 文献级别演化的本质及意义。
3. 文献信息综合应用中“综合”二字的含义。
4. 科技图书与科技期刊之于学术生涯的作用。
5. 信息检索系统构成与信息检索原理的关系。
6. 文本文献检索技术主要有哪些？
7. 布尔逻辑运算符的意义。
8. 截词检索有几种？
9. 文献检索途径主要有哪几种？
10. 如何构建检索策略？

第二章 文献信息检索语言

为了更加科学而有效地汇集、组织、存贮和传播文献信息,人类在自然语言的基础上创建形成了专门用于文献信息存贮与利用的语言系统——信息检索语言。同自然语言一样,它同样由语汇和语法所构成。但其语汇的创设或选择、语法规则的制定或借用,均基于文献信息内容与形式的基本特征、基于人类信息需求的特殊性和多样性、基于文献信息系统的基本原理和核心使命。信息检索语言,奠定了文献信息检索系统构建的语言学基础,构筑了文献信息检索的理论内核,决定了信息检索系统的结构与功能。因此,了解信息检索语言的基本知识,有助于深层次理解检索原理与技术方法。

第一节 文献信息检索语言概述

一、检索语言的概念

文献信息检索语言简称检索语言(retrieval language),是在文献检索领域内用来描述文献特征和表达检索提问的一种专用语言,即根据信息检索需要而创建的统一文献标引用语和检索用语的一种人工语言。它用于检索系统的构建、检索工具的编制和使用,并为检索系统提供统一的、基准的、用于信息交流的一种符号化或语词化的专用语言。检索语言因其使用的场合不同有不同的称呼。例如在文献存储过程中用于文献的标引,称作标引语言;用于文献的索引,则称作索引语言;在文献检索过程中则为检索语言。

检索语言是文献信息检索的重要组成部分,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。因此,检索者有必要学习其中的主要规则、基本原理,减少漏检和误检,提高检索效率。

二、检索语言的构成

为了将文献中使用的自然语言转换成检索时使用的检索语言,并用一定的文字、符号形式予以固定和表达,则需要建立检索词典。检索词典(retrieval thesaurus)是文献标引用语和检索用语的语源和依据性文本。它是对各学科的名词术语、概念、代码、分类号等进行规范化的记录,起着规范控制自然语言的作用。最常见的检索词典是各种分类表和主题词表。无论是何种检索词典,构成检索语言通常应具备3个基本要素:

①有一套用于构词的专用字符。字符是检索语词的具体表现形式,它可以是自然语词中的规范化名词或名词性词组,也可以是具有特定含义的一套数码、字母或代码。

②有一定数量的基本词汇用来表述各种基本概念。基本词汇是组成一部分类表、词表、代码表等全部检索语词标识的总汇,如分类号码的集合就是分类语词的词汇,一个标识(分类号、检索词、代码)就是一个语词。

③有一套专用语法规则来表达由各种复杂概念所构成的概念标识系统。标识是对文献信息特征所做的最简洁的表述。标识系统是对全部标识按一定的逻辑关系编排组合成的有序的

整体。语法是指如何创造和运用那些标识来正确表达文献信息内容和信息需求,以有效实现信息检索的一整套规则。

任何一种信息检索语言,都要采取一定的方法和手段,将它的各种要素按照一定的结构合成一个有机的统一体,以便发挥其最佳功能。

三、检索语言的作用

检索语言是信息检索系统存储与检索共同遵循的一种专用语言,它既是汇集、组织、存储文献的标准,也是检索提问时所利用的手段及工具。文献信息检索包括信息的存储和查找两个过程,它们有着密不可分的关系,互为依存,而它们的实现有赖于检索系统的存在。因此,信息检索的基本原理就是在检索系统的形成和利用过程中使文献信息的存储与查找所采用的检索语言达到一致。

作为文献特征标识与检索提问标识共同使用的检索语言,它规范了文献信息标引人员和检索人员都要用相同的语言来表达同一主题概念的内容,即排除了自然语言中不适合于检索的部分,从而使信息存储和查找二者之间所依据的规则保持一致性,这样才能使文献信息存得进,又取得出,实现了信息检索的全过程。否则,信息检索也就不可能顺利实现,甚至根本不能实现。可见,检索语言是信息标引人员和检索人员之间进行思想交流的媒介,也是人与检索系统之间交流的桥梁,在信息检索过程中起着语言保障的作用,主要表现在:①对文献的各种特征加以标引,是文献信息的组织、存储和检索的基础与前提;②对文献内容相同及相关的信息加以集中或揭示其相关性;③对大量文献信息加以系统化或组织化,形成各种标识系统或索引系统;④便于将标引用语和检索用语进行相符性比较。

四、检索语言的类别

全世界有数以千计的信息检索语言,但任何一种检索语言,都是表达一系列概括文献信息内容的概念及其相互关系的概念标识系统,它们可用于对文献信息的内容进行主题标引、逻辑分类或特定信息的揭示与描述。因此,构成各种检索语言的基本原理是一致的,只是在表达各种概念及其相互关系时和解决对它们提出的共同要求时所采用的方法不同,才形成了不同类型的检索语言,构成了不同的标识系统和索引系统,从而提供了不同的检索途径。

文献信息检索语言有很多类型,可以按不同的方式和标准划分。

1. 按照标识的组合方法划分,可分为先组式语言和后组式语言。

(1) 先组式语言:是文献标识在编表时就固定组配好,或绝大部分已固定组配好的一类检索语言。如体系分类语言、标题词语言均属这种类型。这种语言标识明确,系统性较好,适用于传统的文献单元方式的目录索引,是检索用户比较习惯的形式。

(2) 后组式语言:是文献标识在检索时才组配起来的一种检索语言类型。如单元词语言、叙词语言均属此类。这种语言采用概念分析和综合的原理,适用于标识单元方式的检索系统,可实行多途径、多因素检索乃至精确检索,相当灵活,检索效率较高。但是,标识明确性较差,读者使用也不甚习惯。

2. 按照检索语词规范化程度划分,可分为规范语言和非规范语言。

(1) 规范语言(controlled language):亦称人工语言(artificial language),是人为地对标引用词或检索用词加以控制和规范,使每一个词对应表达一个概念。这些语言经过规范化处理,词和概念之间具有一一对应的关系,排除了自然语言中同义词、多义词、近义词和同形异义词的现象。例如“肿瘤”这一概念在英语中有多个表达方式:cancer, tumor, tumour, carcinoma, neoplasm等,但在规范化语言中,人为规定以neoplasm来表达“肿瘤”这一概念,无论在原始文献中使用哪一个词,只要使用neoplasm一词进行检索,在检索结果中将包括全部含有“肿瘤”概

念的信息。可见,使用规范语言检索既可省略对该概念的全部同义词或近义词的考虑,也可避免多次输入检索词的麻烦和由此出现的误差,保证检索质量和检索效率,但对文献标引人员和检索人员在选词上要求比较严格。分类语言中的分类类目、主题语言中的叙词和标题词都属于规范语言的范畴。

(2) 非规范语言(uncontrolled language): 亦称自然语言(natural language), 是以未经人工控制的、直接从原始文献信息中抽取出的自由词(语词或符号等)作为检索词。这些自由词具有较大的弹性和灵活性,能及时反映最新的概念和规范词难以表达的特定概念,检索者可以自拟词语进行检索。但这类检索语言缺乏对词汇的控制能力,也无法揭示概念之间的关系,存在大量同义词、多义词现象和含义模糊现象。主题语言中的关键词和单元词就属于此类。

3. 按照所描述的文献信息特征划分,可分为文献外表特征检索语言和文献内部特征检索语言。

(1) 文献外表特征检索语言: 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检索语言(系统)。常见的有: ①以文献上记载的书名、刊名、篇名等作为检索标识的文献名称索引系统,如书刊目录等; ②以文献中署名的著者、译者、编者等姓名或团体机构名称作为检索标识的著者索引系统,如著者索引等; ③以文献特有的序号作为检索标识的文献序号索引系统,如专利号索引、科技报告号索引等; ④以文献末尾所附的参考文献或引文的外表特征作为检索标识的引文索引系统,如引文索引等。

(2) 文献内容特征检索语言: 是依据文献所论述的主题、观点、见解和结论以及构成原理作为文献存储的标识和文献检索提问的出发点而设计的检索语言(系统)。主要有以下3种:

1) 分类检索语言: 是把各种概念按所属学科范畴进行分类和分层次系统排列的一种语言体系。具体而言,分类语言是以分类法的形式来体现的,分类法是将代表各种概念的类目用号码来标识,以学科属性为主并加以系统排列来组织文献的体系。它能集中体现学科的系统性,反映事物之间的从属、派生和平行等关系,并从总体到局部分层、分面展开,是一种等级分明的语言。这种语言一般是用分类号(数字或数字与字母组合)作为检索标识来表达各种概念,使同一学科专业文献集中,提供从学科专业角度查找文献信息的途径(详见本章第二节)。

2) 主题检索语言: 是用语词作为检索标识来表达各种概念、并按字顺组织起来的一种检索语言。主题检索语言具有专指性和直接性的特点。根据其表达概念的不同形式又分为标题词语言、单元词语言、关键词语言和叙词语言,其中应用较多的是关键词语言和叙词语言。

标题词语言(heading)是最早使用的检索语言,其使用的词汇不是书名或篇名,而是来自于文献内容特征并是规范化的检索语言。标题词语言属于先组式语言,故灵活性较差。

单元词语言(uniterm)亦称单元词,是指概念上不能再分的最小的语词单位,是一种非规范化语言。例如“肺功能”不是单元词,分割为“肺”、“功能”才是单元词。单元词语言是通过若干单元词的组配来表达复杂的主题概念,具有灵活的组配功能,因此该语言属于后组式检索语言。

关键词语言(keyword)是指从文献题目、文摘或正文中提取出来的并具有实质意义的、能代表文献主题内容的语词。关键词未经规范处理(属于自然语言的范畴)。它在检索工具中常以“关键词索引”(keyword index)作为索引标识系统。关键词语言具有灵活性强、易于掌握、查检方便等特点,尤其广泛应用于计算机检索以及某些最新出现的专业名词术语的查找。但其未经规范化处理,用词不统一,因而有时会出现同一主题内容的文献由于使用不同的关键词而被分散,容易造成漏检,影响查全率。

叙词(descriptor)亦称主题词(subject headings),是指能代表文献实质内容且经过严格规范化的专业名词术语或词组。叙词语言是采用表示单元概念(规范化语词)的组配来对文献主题进行描述的检索语言。两个或两个以上的叙词组配在一起,形成一个新概念,数量不多的叙词可组成许多概念,从而提高了文献标引的专指性和检索的灵活性,叙词语言属于后组式语言,

它是目前使用最广泛的主题语言。该检索语言在检索工具中常以“主题词索引”作为索引标识系统。主题词语言的主要特点有：①它对一个主题概念的同义词、近义词等适当归并，以保证语词与概念的唯一对应，避免多次检索；②采用参照系统揭示非主题词与主题词之间的等同关系以及某些主题词之间的相互关系，以便正确选用检索词；③根据主题词之间的隶属关系，可编制主题词分类索引，从而选择更专指的主题词；④同一篇文献的每个主题词都可以作为检索词，从而提供多个检索入口点，便于查找。基于主题词特点，则需要构建一部供标引和检索使用的主题词表，以保证对主题词语言的正确使用。最常用的医学主题词表是美国国家医学图书馆出版的《医学主题词表》(Medical Subject Headings, MeSH)。

3) 代码检索语言：即各种代码系统，它是对文献所论述事物的某一方面特征，用某种代码系统加以标引和排列的一种检索语言，如美国《化学文摘》的化学物质分子式索引系统。

4. 本体语言 随着计算机技术和网络技术的迅速发展，现代信息组织特别是网络信息组织对信息检索语言提出了新的要求与挑战，依赖于传统文献信息检索语言的信息组织方法已不能完全满足新的信息环境和检索要求，本体(ontology)语言的出现则为这一问题的解决提供了新的思路。

本体是一个哲学的概念，原意指关于存在及其本质和规律的学说，后来被计算机科学领域引入，并赋予了新的内涵。本体能够将某个领域中的各种概念及概念之间的关系显性地、形式化地表现出来。它强调领域中的本质概念，也强调这些概念之间的关联。本体语言是由经过精确定义的可供计算机识别的概念和概念之间的关系构成，是对领域知识的共同理解与描述，它可以更好地用于共享、交流和重复使用。

从本体的概念及其功能特点来看，本体具有与信息检索语言相类似的一些功能特点：①具有与信息检索语言相一致的本质特征，都是对文献内容的概念及相关关系的描述；②信息检索语言的基础也是本体的基础，即本体本身也是基于信息检索语言中的概念逻辑和知识分类的；③具有与信息检索语言同样的职能与作用，本体与信息检索语言一样可以起到对信息的存贮和检索作用，是沟通标引人员与检索人员双方思想的桥梁。可见本体包含或具有与信息检索语言相一致的本质特征、基础、作用和功能。因此，当本体应用于信息检索领域、发挥信息检索语言作用与功能时，可以将本体看作是一种全新的信息检索语言。

本体是网络信息组织的核心概念，本体方法作为一种全新的信息组织方法，它具有传统信息检索语言和信息组织方法所无法比拟的一些功能和特点，作为一种能在语义和知识层次上描述信息系统的概念模型的建模工具，已被广泛应用于知识工程、系统建模、信息处理、数字图书馆、自然语言理解、语义 web 等领域之中，为信息组织特别是网络信息组织带来了许多新的变革。

第二节 文献分类法

一、文献分类法概述

文献分类法是根据文献知识内容所属的学科范畴，分门别类地、系统地揭示和组织文献的一种方法。它是由许多类目根据一定原则组织起来，通过标记符号来代表各级类目并固定其先后次序的分类体系。文献分类法体现在分类表上，分类表是由一系列分类号集合而成。它包括等级体系分类法和组配分类法两种，应用较多的是前者。这是一种直接体现知识分类的等级制概念标识系统，其主要目的是为了满足不同人们按学科知识体系检索文献的要求。图书馆通常都采用文献分类法来编制分类目录和组织馆藏文献的分类排架。

目前，各国均根据自己的文献出版收藏情况，编制了相应的文献分类法，即将所有学科的



图书、期刊或其他资料按其学科内容分成若干个大类，每一大类下分许多小类，每一小类下再分子小类等，最后构成一部比较完善的文献分类体系。从而使每一种文献都可以分到某一个类目下，每一个类目都有一个类号。无论哪种文献分类法，其内容结构一般由以下几个部分组成。

1. 编制说明 主要说明该分类法的编制过程，所依据的编制原则、类目设置和编制的理由，对各类分类问题的技术处理和使用、标引方法等。

2. 类目表 是分类表的主体部分。其主要包括：①大纲，是基本大类的一览表；②简表，是一个基本类目表，它是整个类目表的构架；③详表，即主表，详细列出大小类目、类号和注释；④辅助表，也称复分表，是一组标准目次表，用以对主表中列举的类目进行细分。

3. 索引 即分类法的类目名称索引，按字序排列，以供查询，帮助不熟悉分类法的读者使用。

4. 附录 即收录按类检索时经常要查阅的一些参考资料。

世界上有许多版本的文献分类法。现对国内外目前较有影响力的、常用的分类法作一些简单的介绍。

二、常用文献分类法

1. 《中国图书馆分类法》 简称《中图法》，是目前我国使用最广泛的一种等级体系分类法。它不仅应用于各类型图书馆的藏书排架和组织目录体系，绝大多数的文献检索系统也都是按《中图法》的分类体系编制和提供服务的。该分类法由北京图书馆、中国科学技术情报所等单位共同编制完成的，于1974年出版，并经过多次修订与再版，目前已修订至第五版。

《中图法》由五大部类、22个大类、8个通用复分表、66多个专类复分表、51 881条类目组成了一个比较完整的分类体系。为体现整部分类法的指导思想，把“马克思主义、列宁主义、毛泽东思想、邓小平理论”作为一个基本类列于首位，“哲学、宗教”“社会科学”“自然科学”分别按其知识的逻辑关系列为第二、三、四位，最后一个部类为“综合性图书”，收录一些内容涉及面广、类无专属的文献。

《中图法》的标记制度采用英文字母与阿拉伯数字相结合的混合小数层累制，除“工业技术”类外，其余各大类均用1个英文大写字母表示一级类目名，并以字母的顺序反映大类的顺序。根据各学科的内容需要，大类下再分各级子类目，各子类目用阿拉伯数字标记分类号。数字的编号顺序，采用小数层累标记制。号码的位数代表相应类目的分类等级。当一个分类号的数字超过三位时，加上小圆点“.”便于区分。以上基本构成了《中图法》的分类体系(表2-1)。

表 2-1 《中图法》的 22 个基本大类

A 马克思主义、列宁主义、毛泽东思想、邓小平理论	N 自然科学总论
B 哲学、宗教	O 数理科学和化学
C 社会科学总论	P 天文学、地球科学
D 政治、法律	Q 生物科学
E 军事	R 医药、卫生
F 经济	S 农业科学
G 文化、科学、教育、体育	T 工业技术
H 语言、文字	U 交通运输
I 文学	V 航空、航天
J 艺术	X 环境科学、安全科学
K 历史、地理	Z 综合性图书

在“R医药、卫生”这个一级类目下又分出17个二级类目(图2-1)。

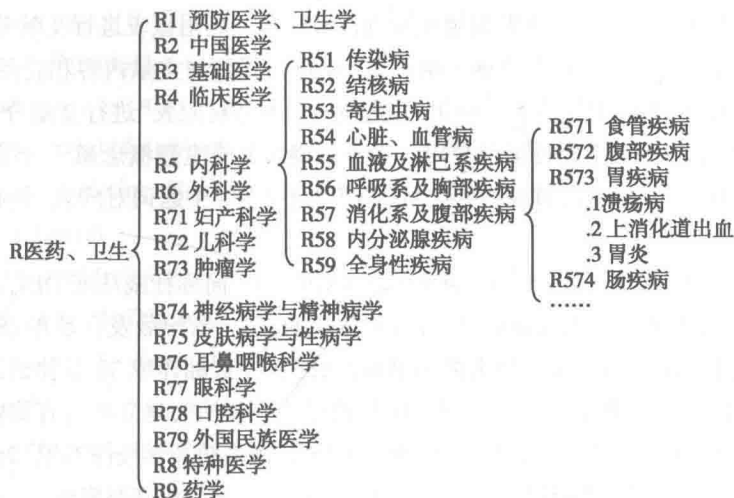


图 2-1 “R 医药、卫生”的主要类目

有关不同学科的同类图书(如工具书、教科书等)亦可按上述分类查找。为了与一般图书相区别,有时在分类号末尾增加一个带有统一连字符的复分号,形成新的更专指的分号。例如《中国内科年鉴》的分类号是“R5-54”,其中“R5”是“内科学”的分类号,“-54”是复分号,表示“年鉴、年刊”。常用的复分号有:

- 41 教学计划、教学大纲、课程
- 42 教学方法、教学参考书
- 43 教材
- 44 习题、试题及解答
- 45 教学试验、实习
- 49 普及读物
- 5 丛书、文集、连续性出版物
- 51 丛书(汇刻书)、文库
- 52 全集、选集
- 53 论文集(会议录、毕业论文、学位论文等)
- 54 年鉴、年刊
- 6 参考工具书
- 61 名词术语、词典、百科全书(类书)
- 62 手册、名录、指南、一览表、年表
- 63 产品目录、样本、说明书
- 64 表解、图表、图册、谱录、公式、地图
- 65 条例、规程、标准
- 66 统计资料
- 67 参考资料
- [7] 文献检索工具

此外,为了能同时满足文献标引和文献检索的需要,我国在《中图法》和《汉语主题词表》的基础上又编制了《中国分类主题词表》,它是两者兼容的一体化信息检索语言,共收录分类法类目 5 万多个,主题词及主题词串 21 万余条。《中国分类主题词表》分为 2 卷,第 1 卷为“分类号-主题词对应表”(2 册),第 2 卷为“主题词-分类号对应表”(4 册)。“分类号-主题词对应表”中的每条对应款目分为左、右栏两部分,左栏是分类号、类名和类目注释,其格式与《中图法》基

本相同;右栏是对应的主题词和主题词组配标题以及注释。利用该表进行文献分类标引或检索时,主要是利用左栏,由大类到小类逐步缩小范围,从而找到与文献内容相符的类目。这等于使用《中图法》进行分类标引或检索。利用“主题词-分类号对应表”进行文献分类标引或检索,实际上是把它作为《中图法》类目索引使用。当难以确定某个主题概念属于分类体系中的哪个类目时,可从该表入手,先查出对应的分类号,再查“分类号-主题词对应表”进行核对,从而确定具体分类号。

2. 杜威十进制分类法(*Dewey Decimal Classification*) 简称杜威法或 DDC,它是由美国图书馆学家麦维尔·杜威(Melvil Dewey)于 1873 年创建的,第一次出版发行是在 1876 年。该分类法是目前全世界使用最广泛、影响最大的图书资料分类法,被翻译成 30 多种语言,用于 138 多个国家的 200 000 多个图书馆。对全世界各国其他图书分类法的建立亦具有重要影响,许多分类法均是借鉴于杜威法产生的。在美国,有 95% 的公共图书馆及学校图书馆、25% 的学院及大学图书馆及 20% 的专门图书馆使用 DDC。此外,DDC 能用来组织网络资源。

2011 年 OCLC 正式出版 DDC 第 23 版的印刷版,即 DDC-23。DDC 用传统的学科体系进行分类,以 10 个主要的学科大类(main classes)来涵括所有的知识体系,每个大类下细分 10 类(divisions),每个类下又再分成 10 小类(sections)。DDC 中每个学科都会用特定范围的数字来表示,它的十个大类(main classes)分别是:

- 000 Computer science, information & general works 计算机科学,信息与综合类
- 100 Philosophy & psychology 哲学与心理学
- 200 Religion 宗教
- 300 Social sciences 社会科学
- 400 Language 语言
- 500 Sciences 自然科学
- 600 Technology 应用科学(包括医学类)
- 700 Arts & recreation 艺术及娱乐
- 800 Literature 文学
- 900 History & geography 历史与地理

关于《杜威十进制分类法》更详细的信息参见其网址: <http://www.oclc.org/dewey/resources/summaries/Deweysummaries.pdf>。

3. 美国国家医学图书馆分类法(*National Library of Medicine Classification*, NLMC) 是专门针对健康科学而设的文献分类法。自 1951 年正式发行第 1 版以来,经过多次修订,现为 2000 年出版的第 5 版修订版。2002 年编制电子版,其分类法提供索引中的类号与总表类号的超链接,并且索引用语可直接连接至 MeSH Browser 的标题。自 2002 年起, NLMC 每年仅以电子版的形式发行。NLMC 的网址是 <http://www.nlm.nih.gov/class>。

NLMC 采用字母数字混合号码标记类目。它与《美国国会图书馆分类法》(LCC)关系密切, NLMC 利用 LCC 中的空号 QS~QZ 表示基础科学(preclinical science)和医学(medicine and related subjects)。另外,美国国家医学图书馆也极少使用 LCC 中的 K(Law),而将有关医学的法律文献根据相关主题各入其类。以下为该分类法的简表:

Preclinical Sciences	基础医学
QS Human Anatomy	人体解剖学
QT Physiology	生理学
QU Biochemistry, Cell Biology and Genetics	生物化学、细胞生物学和遗传学
QV Pharmacology	药理学
QW Microbiology and Immunology	微生物学与免疫学

QX	Parasitology	寄生虫学
QY	Clinical Pathology	临床病理学
QZ	Pathology	病理学
Medicine and Related Subjects		医学及相关学科
W	General Medicine, Health Professions	医药、卫生行业
WA	Public Health	公共卫生
WB	Practice of Medicine	医学总论
WC	Communicable Diseases	传染病
WD	Disorders of Systemic, Metabolic or Environmental Origin, etc.	系统性疾病、代谢疾病或环境诱发疾病等
WE	Musculoskeletal System	肌肉骨骼系统
WF	Respiratory System	呼吸系统
WG	Cardiovascular System	心脏与血管系统
WH	Hemic and Lymphatic System	血液与淋巴系统
WI	Digestive System	消化系统
WJ	Urogenital System	泌尿生殖系统
WK	Endocrine System	内分泌系统
WL	Nervous System	神经系统
WM	Psychiatry	精神病学
WN	Radiology, Diagnostic Imaging	放射学、诊断成像
WO	Surgery	外科学
WP	Gynecology	妇科学
WQ	Obstetrics	产科学
WR	Dermatology	皮肤科学
WS	Pediatrics	儿科学
WT	Geriatrics, Chronic Disease	老年医学, 慢性病学
WU	Dentistry, Oral Surgery	牙科学, 口腔外科
WV	Otolaryngology	耳鼻喉科学
WW	Ophthalmology	眼科学
WX	Hospitals and other Health Facilities	医院及其医疗设备
WY	Nursing	护理学
WZ	History of Medicine, Medical Miscellany	医学史及其他

第三节 国际疾病分类法

一、国际疾病分类法概述

疾病是指人体在一定条件下,由致病因素所引起的一种复杂而有一定表现形式的病理过程。疾病的种类很多,且患病部位、致病因素、病理改变和表现的症状也呈多样性和复杂性。疾病分类就是根据疾病的本质、特点,将疾病分门别类地划分与归类,并使其成为一个有序的组合系统。对疾病进行分类是深入认识各种疾病的必然结果。因为它不仅对医药卫生单位开展医疗、科研、教学,以及搞好病案管理、资料信息和卫生统计等工作是必不可缺的,而且还能

中的分布特征及其发生、变化的规律性,制定预防保健措施,增进人类健康。

国际疾病分类法(*International Classification of Diseases, ICD*)就是按照既定疾病分类标准将各种疾病名称归入相应类目的一种系统。它是对疾病现象进行数量研究和在国际间进行医学科研学术交流的基础工具。其目的是为了对不同国家或地区以及在不同时间内收集到的死亡和疾病数据进行系统地记录、分析、解释和比较。同时它还把疾病诊断和其他健康问题的词句转换成字母数字编码,从而易于对数据进行储存、检索和分析。

ICD 是一部已有百年历史、经过多次讨论和修改的国际上统一使用的疾病分类法。国际疾病分类方法源于 1853 年在布鲁塞尔召开的第一次国际统计学大会,会议提出制定全世界范围使用的统一疾病名称。1855 年在巴黎召开的第二次国际统计学大会上通过了威廉·法尔提出的 138 个项目的疾病分类表。这个分类表虽然未获得举世公认,但其疾病分类的一个重大原则——把全身性疾病和局限于某器官解剖部位的疾病区别开来的原则,却一直保存下来。1893 年在芝加哥召开的国际统计学研究大会上,通过了《贝蒂荣死亡原因分类法》,它获得普遍赞许,并被许多国家和地区所采用。1900 年法国政府在巴黎召开了对《贝蒂荣死亡原因分类法》修订的国际会议,会上通过了《国际死亡原因分类法》,即国际疾病分类法的第一个蓝本。其后每隔 10 年左右召开一次国际修订会议。1948 年世界卫生组织(WHO)接管国际疾病分类工作,并主持召开第六次以后的国际修订会议。自从世界卫生组织接管以后,国际疾病分类工作进入一个新的发展时期。它从一个局限的死亡原因分类体系扩大成为全面的疾病分类体系,除疾病死亡统计外,开始应用于疾病管理等方面,并自三位数字编码系统向小数点下发展。

目前使用的国际疾病分类法是由世界卫生组织于 1993 年编撰的第十次修订版本(ICD-10),改称为《疾病及有关健康问题国际统计分类法》(*The International Statistical Classification of Diseases and Related Health Problems*),为保持其连续性,简称仍沿用“国际疾病分类”(ICD)。但其使用范围不仅局限于疾病和损伤分类的范畴,还扩展到对其他健康问题进行分类。

二、学习国际疾病分类法的意义

在 WHO 的倡导和推动下,ICD 已为大多数成员国所接受,成为世界性的疾病、损伤和死亡原因分类的统一标准化工具。它对促进世界卫生保健事业,掌握医疗卫生工作动态以及卫生情报的国际交流,起着极其重要的作用。我国是 WHO 的成员国,有义务按照 ICD 系统进行疾病、死亡原因的编码和分类,并向 WHO 呈报我国的卫生统计信息。我国于 1987 年起正式使用 ICD-9 进行疾病和死亡原因的统计分类。国家技术监督局于 1993 年 5 月发布了等效采用 ICD-9 编制的“疾病分类与代码”国家标准。为此,我国卫生部要求县级及以上医院采用 ICD 编制医院出院病人疾病分类统计报告,并颁发了全国统一使用的、注有 ICD 编码的《医院病案首页》,首页中很多项目内容(如各种诊断、手术名称、损伤、根本死因等)都需要临床医生参照 ICD 来正确书写。因此,各级临床医生了解国际疾病分类的使用规则是非常必要的。此外,国际疾病分类法中还记载了许多新的疾病、损伤、药物及化学物质、手术名称等信息,这对拓宽医学知识及科研课题的开展也有一定作用。

三、国际疾病分类法的主体结构

ICD-10 印刷版本由三卷(独立成册)组成。第一卷是 ICD 编码的主要内容类目表;第二卷是使用指导手册;第三卷是分类的字顺索引。

第一卷包括三位数类目表、内容类目表和四位数亚目、肿瘤形态学分类、死亡和疾病的特殊类目表、定义和命名条例。其中三位数类目表为核心分类表,是向世界卫生组织死亡率数据库提交报告和一般国际比较用的强制性编码水平。在国际水平提交报告时,并不强制使用四位数亚目。三位数类目表(list of three-character categories)共有 21 章(大类)(表 2-2)。

表 2-2 ICD-10 的三位数类目表

章	类目名称	类目编号
一	某些传染病和寄生虫病	A00~B99
二	肿瘤	C00~D48
三	血液及造血器官疾病和某些涉及免疫系统的疾患	D50~D89
四	内分泌, 营养和代谢疾病	E00~E90
五	精神和行为疾患	F00~F99
六	神经系统疾病	G00~G99
七	眼和附器疾病	H00~H59
八	耳和乳突疾病	H60~H95
九	循环系统疾病	I00~I99
十	呼吸系统疾病	J00~J99
十一	消化系统疾病	K00~K93
十二	皮肤和皮下组织疾病	L00~L99
十三	肌肉骨骼系统和结缔组织疾病	M00~M99
十四	泌尿生殖系统疾病	N00~N99
十五	妊娠、分娩和产褥期	O00~O99
十六	起源于围产期的某些情况	P00~P96
十七	先天畸形、变形和染色体异常	Q00~Q99
十八	症状、体征和临床与实验室异常所见, 不可归类在他处者	R00~R99
十九	损伤、中毒和外因的某些其他后果	S00~T98
二十	疾病和死亡的外因	V01~Y98
二十一	影响健康状态和与保健机构接触的因素	Z00~Z99

每章又列出了数目不等的各节标题, 每节标题包括若干个类目, 每一类目中都有疾病名称及相应的三位数编码(第一位为英文字母, 第二、三为数字; 每一字母都与特定章节有关)。如“高血压病”是“循环系统疾病”(I00~I99)中的一节, 其标题下面的类目及编码是:

Hypertensive diseases (I10~I15)

- | | | |
|-----|--------------------------------------|-------------|
| I10 | Essential (primary) hypertension | 特发性(原发性)高血压 |
| I11 | Hypertensive heart disease | 高血压性心脏病 |
| I12 | Hypertensive renal disease | 高血压性肾脏病 |
| I13 | Hypertensive heart and renal disease | 高血压性心脏和肾脏病 |
| I15 | Secondary hypertension | 继发性高血压 |

内容类目表和四位数亚目 (tabular list of inclusions and four character subcategories) 列出了各章和节的标题及其四位数水平级别的全部类目, 实际上它是三位数类目表的详细类目表。在表中所有的四位数字以上的均用“0~9”来表示, 并用“.”和前面的三位数分开, 表明亚目。其中用“0~7”来表示已列出的疾病, 8 表示“其他”未列出的疾病类目, 9 表示“未特指”疾病类目。如“继发性高血压”的四位数亚目是:

- | | | |
|---|---|---------------|
| I15 | Secondary hypertension | 继发性高血压 |
| Excludes: involving vessels of: 不包括: 累及……的血管: | | |
| | • brain (I60~I69) | • 脑 (I60~I69) |
| | • eye (H35.0) | • 眼 (H35.0) |
| I15.0 | Renovascular hypertension | 肾血管高血压 |
| I15.1 | Hypertension secondary to other renal disorders | 继发于其他肾疾患的高血压 |

- I15.2 Hypertension secondary to endocrine disorders 继发于内分泌疾患的高血压
 I15.8 Other secondary hypertension 其他继发性高血压
 I15.9 Secondary hypertension, unspecified 继发性高血压, 未特指

第二卷是指导手册, 包括使用说明、编码的规则和指导、统计报告以及 ICD 的发展史。

第三卷是第一卷分类的字顺索引, 包括三部分: 疾病和损伤性质的字顺索引、损伤的外部原因索引、药物和化学制剂索引。整个索引有关术语和编码 6.3 万余条(其中相当一部分是在第一卷没有出现的术语)。各部分索引均按主导词的字顺排列, 其下是不同水平的修饰词或限定词。一个完整的索引术语的构成是对主导词与修饰词或限定词采用错行缩格的形式, 其后是该术语的编码。因此, 通常一个完整的索引术语可能由几行构成。如“继发性高血压”在索引中的著录内容及格式是:

高血压 I10

.....

-secondary I15.9 继发性

--speci.ed NEC I15.8 特指的

--due to 由于

---endocrine disorders I15.2 内分泌疾患

---renovascular disorders I15.0 肾血管疾患

---renal disorders NEC I15.1 肾脏疾患

----arterial I15.0 动脉

----pheochromocytoma I15.2 嗜铬细胞瘤

在使用 ICD-10 进行编码时, 首先根据诊断对象来确定主导词, 选择主导词正确与否关系到能否找到准确的疾病编码。其次根据主导词及修饰或限定部分的具体要求, 在第三卷的有关索引中查找适当的编码; 最后在第一卷的类目中核对编码。

第四节 国际系统医学术语集

一、发展沿革及应用

1. 发展沿革 国际系统医学术语集(Systematized Nomenclature of Medicine-Clinical Terms, SNOMED CT, <http://www.ihtsdo.org/snomed-ct/>)是一个能提供医学信息存储、提取与交换的医学标准术语系统。

SNOMED CT 的前身是以病理学术语为核心制定的《系统病理学术语集》(Systematized Nomenclature of Pathology, SNOP), 于 1965 年首版, 是一个包括术语及其相关联编码的四轴(axis)系统, 旨在为病理学家提供一个医学数据存储及检索的工具。

1974 年 SNOP 更名为 SNOMED, 此时其应用范围已超出病理学范畴。1979 年 SNOMED 2 发行, 联合国国际医学科学组织病理学分会认证 SNOMED 2 为通用的临床和病理解剖学术语, SNOMED 也得到美国临床病理家学会和英国皇家病理家学院的支持。1993 年, 美国临床病理家学会和美国国家医学图书馆合作将 SNOMED 整合入 UMLS(详见本章第五节), 由此 SNOMED 成为 UMLS 的一个重要子集。同年 9 月 SNOMED 3 问世, 为医学诊断和治疗提供了多维化、结构化的术语检索, 成为 SNOMED 发展史上第一个被世界范围广泛接受的版本, 并被翻译成多种语言出版。1996 年, 美国临床病理家学会与美国放射学学院及美国国家设备制造协会合作, 在他们的《医学数字成像与通讯》(Digital Imaging and Communications in Medicine, DICOM)标准中开发了一个属于 SNOMED 的子集, 有利于与医疗影像相关的信息进行传

递。1997 年, SNOMED 与《观察结果标识逻辑命名与编码系统》(Logical Observation Identifiers Names And Codes, LOINC)进行了相关临床参考术语的交叉映射,由 LOINC 提供的与实验室检查相关的具体术语名称被映射到 SNOMED 内部。

2000 年又进一步推出 SNOMED RT (SNOMED Reference Terminology), 该版本通过逻辑描述、关系型数据库结构等计算机技术来记录卫生保健、疾病分期、病理生理学研究、治疗及治疗结果所需要的语义学上统一的临床术语与医学信息编码,并提供了电子病历的底层框架支持,使用户可以通过任何一种指标去查询某个具体案例及与 LOINC、ICD 的映射代码。

2002 年美国临床病理家学会与英国国家卫生局合作推出 SNOMED CT, 将 SNOMED RT 与英国的 CTV3 (Read Codes Version 3) 两种术语集合并,成为当时包含内容最为广泛的国际性临床参考术语集,为整个医疗行业的国际交流提供了一个跨越语言和地域的平台。2002 年和 2003 年先后推出了 SNOMED CT 的西班牙文版和德文版。2005 年成立了非营利性国际机构 IHTSDO (International Health Terminology Standards Development Organization) 以更有效地管理 SNOMED CT。目前可以获取包括美国英语、英国英语、西班牙语和丹麦语种的 SNOMED CT 版本,并正在将 SNOMED CT 翻译成法语、瑞典语、立陶宛语和其他语种的版本。其发展沿革见图 2-2。

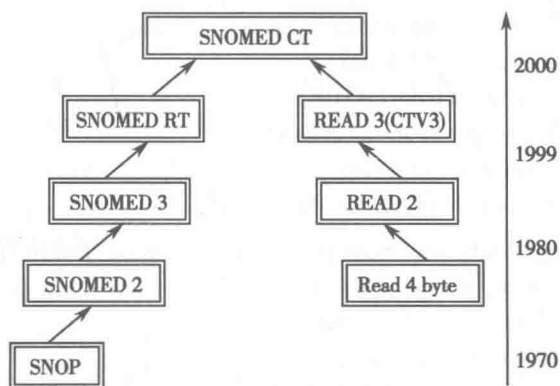


图 2-2 SNOMED CT 发展沿革

2. 应用 SNOMED CT 可应用于电子病历、ICU 监测、临床决策支持、医疗研究、临床实验、计算机处理的医嘱、疾病监测、图像指数、药物管理和公众健康信息服务等以提高处理各项临床医学事务的能力。

目前 SNOMED CT 已得到很多国际标准化组织如 ISO、ANSI 等重要机构的认证,得到一些专业团体如美国护士协会 (ANA)、美国外科医师学会 (ACOS) 等的认证。美国食品与药品管理局 (FDA) 已计划采用 SNOMED CT 作为标准计算机化医疗词汇系统,并将位于处方药标识中强调区的重要术语进行电子化编码,从而允许医学人员在全国范围内电子化地存取数据,更方便有效地共享关键的健康与治疗信息。2002 年由中医科学院承担的北京市重大项目“中医药防治重大疾病临床个体诊疗评价体系的研究”的临床信息采集系统,采用 SNOMED CT 作为医学信息标准编码,一方面为采集患者的临床数据提供了极大的灵活性和可扩展性,另一方面也为将来的数据交流与共享,以及进行医疗科研、建立数据仓库提供有力保障。

二、概念体系

2009 年版 SNOMED CT 包括 400 000 余个概念 (concepts), 这些概念按照从一般到具体的等级结构组织起来,把临床数据记录并汇总为一个更普遍的概念,每个概念有一个唯一的标识符 (concept ID) 和完整的指定名称 (fully specified name, FSN)。与 FSN 相对应的还有其他两

种概念描述形式, 优先词(preferred term)和同义词(synonym)。以慢性阻塞性肺疾病(chronic obstructive lung disease, COPD)为例说明概念体系中概念的相关组成部分(表 2-3)。

表 2-3 概念体系中概念的相关组成部分

基本成分	实例	注释
concept ID	concept ID: 13645005	
description ID	description ID: 475423015	每个 FSN, 优先词和同义词都有唯一的标识符
fully specified name (FSN)	chronic obstructive lung disease (disorder) concept ID: 13645005 description ID: 739041012	每个概念均有唯一的一个 FSN
preferred term	chronic obstructive lung disease concept ID: 13645005 Description ID: 23287019	每个概念均有一个优先词, 它是临床医生最常用的概念, 但它不是唯一的
synonym	CAFL-chronic airflow limitation concept ID: 13645005 description ID: 475425010 COPD-chronic obstructive pulmonary disease concept ID: 13645005 description ID: 475431013 CAO: chronic airway obstruction concept ID: 13645005 description ID: 475419011	同义词有相同的 concept ID, 但每个同义词有单独的 description ID
hierarchy/sub-hierarchy	clinical finding (finding) disease (disorder) chronic disease (disorder) chronic anemia (disorder) simple chronic anemia (disorder) organism (organism)	所有概念分为 19 个顶级概念体系

SNOMED CT 概念之间约有 1 360 000 种连接或语义关系, 提供概念的定义及其他特征。其中一种连接类型是“IS_A”关系, 用来定义等级结构中概念的位置, 例如: chronic obstructive lung disease IS_A chronic disease of respiratory system (慢性阻塞性肺疾病 IS_A 慢性呼吸系统疾病)。

SNOMED CT 收录的海量词汇, 通过对概念的划分, 形成 19 个顶级概念体系, 每个顶级概念再分类细化形成包含多层子系统的树状结构, 这 19 个概念如下:

1. 临床表现 (clinical finding) 用来描述临床观察、评估或判断的结果, 包括正常和异常的临床情况。如 clear sputum (finding)——清晰的痰液 (表现)。
2. 操作 (procedure) 用来描述在提供医疗保健服务中执行的多种行为, 这个概念层级体系用来描述的行为较为宽泛, 包括但不局限于侵入性操作、药品使用、影像操作、教育性操作和管理性操作。如 appendectomy (procedure)——阑尾切除术 (操作)。
3. 观察对象 (observable entity) 描述一个问题或一种操作, 并且这个问题或操作能够形成其答案或结果。如 left ventricular end-diastolic pressure (observable entity)——左心室舒张末端血压 (观察对象) 等同于问题“左心室舒张末端血压是多少”, 或者“已测量的左心室舒张末端血压是多少”。
4. 身体结构 (body structure) 包括了正常及异常的解剖结构, “正常解剖结构”概念用来指

定某种疾病或者操作所涉及的身体部位。如 *uterine structure (body structure)*——子宫结构(身体结构)。

5. 有机体(organism) 收录对人类医学和兽医学有重要意义的有机体,在公共卫生报告中十分重要,对临床决策支持系统中进行基于证据的传染性疾病相关草案的制定工作也有重要意义。如 *Bacillus anthracis (organism)*——炭疽杆菌(有机体)。

6. 物质(substance) 用于记录医药产品、食物和化学过敏原、副作用、毒性或中毒等信息以及医生和护理医嘱中相关的活性化学成分。如 *insulin (substance)*——胰岛素(物质)。

7. 药物/生物产品(pharmaceutical/biologic product) 从“物质”中独立出来,将其设为顶级概念目的在于更为明确地区分属于概念分类“产品”的药物产品与其属于概念分类“物质”的化学成分。此概念分类收录的概念表达了多种层级,以满足对多种用途需要的支撑,如计算机化医嘱录入系统(CPOE)、电子处方、决策支持和处方管理等。此概念分类下的所有概念,不论其属于哪一个层级,均带有一个 FSN 的标记即(product),如 *sex hormone product (product)*——性激素产品(产品)。

8. 标本(specimen) 用来描述那些通常从患者处获得,用以检查或分析的实体。标本类概念可以被指定其特殊属性的连接词定义,如标本获得处正常或异常的身体结构、用来采集标本的操作、标本来源以及标本所含的物质等。如 *Calculus specimen (specimen)*——结石标本(标本)。

9. 物理性物体(physical object) 包括了自然和人造的物体,其用途之一是建立使用某些设备的操作类概念(例如导管插入术)的概念模型。如 *Artificial kidney, device (physical object)*——人工肾脏,设备(物理性物体)。

10. 物理力(physical force) 主要作用是描述在一种伤害机制中起作用的物理力。如 *Spontaneous combustion (physical force)*——自燃(物理力)。

11. 事件(event) 包含了能够表达事情发生的概念,其中不包括操作/干预的概念。如 *Earthquake (event)*——地震(事件)。

12. 环境或地理位置(environment or geographic location) 包括了环境的类型及已经命名的地理定位如国家、州和地区的概念。如 *Rehabilitation department (environment)*——康复科(环境)。

13. 社会环境(social context) 含有对卫生保健有意义的社会性条件和形势的概念,内容包括诸如家庭状况、经济状况、人种和宗教传统、生活方式及职业等方面,表达了能够影响到患者健康和治疗的 社会方面的问题。如 *bank clerk (occupation)*——银行办事员(职业)。

14. 分期与分度(staging and scales) 包括用来命名评估的分度以及用来命名肿瘤分期的系统。如 *Dukes staging system (tumor staging)*——Duke 分期系统(肿瘤分期)。

15. 详细病历(situation with explicit context) 表示概念定义的临床部分,包括阳性体征、阴性体征、现病史、既往史以及家族史例如:内镜检查史(*endoscopy arranged*)、心肌梗死病史(*past history of myocardial infarction*)和青光眼家族史(*family history of glaucoma*)。

16. 限定值(qualifier value) 包含了一些能够被用来作为 SNOMED CT 属性关系值的概念,而且这些概念不会在 SNOMED CT 的其他概念分类中出现。如 *mild (qualifier value)*——温和的(限定值)。

17. 特殊概念(special concept) 属于“特殊概念”的一个亚类“inactive concept”(静止的概念),是术语集中所有已经被停止使用概念的上位概念,同时,这些已停止使用的概念能够指向仍在术语集中使用的概念。

18. 记录人工制品(record artifact) 是一种由一个人或者多个人创造的实体,用来为他人提供关于事件或事态的信息。总体来说记录是虚拟的,也就是说它独立于其特定的物理性实例,并由其多种信息要素组成,这里的信息要素通常是文字、短语、句子,也有数字、图表和其他

信息要素。“记录人工制品”不一定是完整的报告或记录,它们可以是更大记录的组成部分。例如一份完全的健康记录是一个“记录人工制品”,同时它可能包含以独立的文档或报告形式存在的其他“记录人工制品”,而相应的,这些记录中也还可以含有更为细小的记录人工制品例如记录的片段甚至是片段头部。

19. SNOMED CT 模型元数据(SNOMED CT model component) 包含支持 SNOMED CT 的技术元数据。

由 19 个顶级概念设定顺序我们可以看出 SNOMED CT 的设计思路,即按照临床实践需要,以医学实际的诊疗过程为指导进行概念梳理,以最大限度方便临床医师的实际应用为原则。另外,SNOMED CT 自身开放性的构架形式可以使新生的医学概念、医疗信息术语均可以通过 SNOMED CT 分类方法迅速添加到术语集中,并按照 SNOMED CT 编码原则实现数字化来满足医学信息学高速发展的需要。

第五节 医学主题词表

一、概 述

《医学主题词表》(MeSH)是美国国家医学图书馆(NLM)研制的用于标引、编目和检索生物医学文献的英文受控词表。它是对生物医学文献进行主题标引以及检索生物医学文献数据库的指导性工具,对提高查全率及查准率具有十分重要的意义。

目前,NLM 使用 MeSH 表为 MEDLINE/PubMed 数据库的 5200 多种世界一流的生物医学期刊进行标引。同时,也对 NLM 包含图书目录、文档和视听资料的数据库进行标引。通常数据库中每条书目信息是和一组描述其内容的 MeSH 词相连,因此可以使用这些 MeSH 词汇来查找这些特定主题的文献。另外,MeSH 表在全世界其他多个国家也得到了广泛的应用,如中国医学科学院医学信息研究所等机构将英文版的 MeSH 表翻译成了中文,并在中国生物医学文献数据库中提供 MeSH 表的中文及英文的电子版,便于中文文献的主题标引和检索。

NLM 提供 4 种方式联机免费获取其电子版及相关信息:① MeSH Browser;② UMLS(Unified Medical Language System)一体化医学语言系统;③ MeSH 网站(<http://www.nlm.nih.gov/mesh>),包括 MeSH 表的全部内容及 MeSH 相关信息;④ MeSH databases,为用户检索 MEDLINE/PubMed 提供帮助。

二、概念体系

MeSH 表的概念体系由主题词、限定词、补充概念和款目词组成。其中,款目词通过“见参照”反映在主题词记录中。

1. 主题词(main headings) 是用于描述主题事物或内容的,唯一具有独立检索意义的经过规范化处理的名词术语。主题词随着医学科学的发展不断增删、调整,以便及时反映医学科学的最新发展、新主题和新事物。

MeSH 表的选词范围包括生物医学文献中能表达与医学或生命科学有关的概念并具有检索意义的常用词汇。词汇以名词为主,可数名词多采用复数形式,不可数名词或表示抽象概念的名词采用单数形式。主题词可以是单个词,如 lung;也可以是词组,如 lung neoplasms。词组形式的主题词一般按自然语言的顺序,但当一组主题词具有某些相同的概念时,采用倒置形式,把同一概念的词、概念中的核心词排在前,修饰、限定的词放在后,如:

anemia, aplastic

anemia, dyserythropoietic, congenital

anemia, hemolytic
anemia, hemolytic, autoimmune

2. 限定词 (qualifiers) 又称副主题词 (subheadings), 是对主题词所探讨的某一方面内容加以限定的词, 其作用是增强主题词的专指性, 通常用组配符“/”与主题词一起使用。例如: hypertension (高血压)/etiology (病因学)。限定词的数量及其可组配主题词的范围有严格规定, 现有 83 个, 每个限定词可以组配的主题词范围详见附录 2。

3. 补充概念 (supplement concept) 又称补充化学物质名称, 用于标引 MEDLINE 中出现的化学物质和毒品等, 在 PubMed 中可用化学物质名称字段 [Nm] 进行检索的概念。

4. 款目词 (entry terms) 亦称入口词, 是指某些主题词的同义词或相关词, 其作用是将自由词引见到主题词。但在机检时可独立使用。

5. 树状结构 (tree structure) 树状结构又称范畴表, 是将全部主题词按其词义范畴和学科属性编排而成的一个分类体系表。该表共设有 16 个大类, 每个大类又细分出 100 多个二级类目 (附录 1)。二级类目下又层层划分, 逐级展开, 像大树的干、枝、叶一样, 使同一概念范围的主题词呈族系展示它们之间的并列、隶属等语义关系。每个主题词有一个相应的树状结构号 (字母或字母与数字的组合), 其中隶属等级的主题词用逐级缩格排列, 同一级主题词按字顺排列, 最多可达 11 级, 形成主题词的树状结构体系。鉴于学科的分化与交叉, 许多的概念可以分属在不同的大类之下, 因而也就会有一个以上的树状结构号, 如“Breast Neoplasms”既属于肿瘤类主题词又属于皮肤与结缔组织疾病, 因此有两个树状结构号 (表 2-4、图 2-3)。

表 2-4 “Breast Neoplasms”的 MeSH 主题词记录

MeSH Heading	Breast Neoplasms
Tree Number	C04.588.180
Tree Number	C17.800.090.500
Annotation	human only; BREAST NEOPLASMS, MALE is also available; for animal use MAMMARY NEOPLASMS, ANIMAL or MAMMARY NEOPLASMS, EXPERIMENTAL: Manual 24.5+, 24.6+; coordinate IM with histological type of neoplasm (IM)
Scope Note	Tumors or cancer of the human BREAST.
Entry Term	Breast Cancer
Entry Term	Breast Carcinoma
Entry Term	Breast Tumors
Entry Term	Cancer of Breast
Entry Term	Cancer of the Breast
Entry Term	Human Mammary Carcinoma
Entry Term	Malignant Neoplasm of Breast
Entry Term	Malignant Tumor of Breast
Entry Term	Mammary Cancer
Entry Term	Mammary Carcinoma, Human
Entry Term	Mammary Neoplasm, Human
Entry Term	Mammary Neoplasms, Human
Entry Term	Neoplasms, Breast
Entry Term	Tumors, Breast
Allowable	BL BS CF CH CI CL CN CO DH DI DT EC EH EM EN EP ET GE HI IM ME MI MO NU
Qualifiers	PA PC PP PS PX RA RH RI RT SC SE SU TH UL UR US VE VI
Entry Version	BREAST NEOPL
Date of Entry	19990101
Unique ID	D001943

主题词的树状结构体系有三个作用：①提供主题词等级关系的完整显示，有利于用户选词扩大或缩小检索范围；②在计算机检索系统中，可以实现自动扩检，即将某词以下的所有主题词进行检索，满足族性检索要求；③通过上位词、下位词及同位词的显示，进一步明确词间关系及词义。

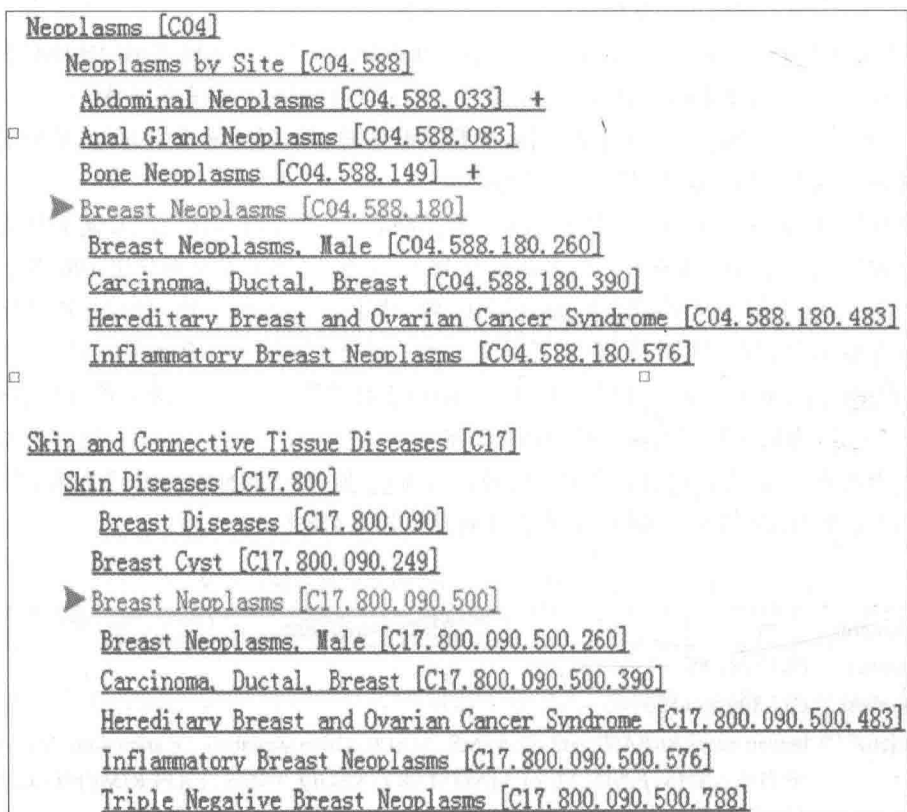


图 2-3 “Breast Neoplasms” MeSH Tree Structures

三、MeSH Browser

MeSH Browser (医学主题词浏览器, <http://www.nlm.nih.gov/mesh/MBrowser.html>) 是 MeSH 表的网络版, 旨在帮助用户迅速查询相关主题词并显示主题词的等级结构体系。

MeSH Browser 提供了两种查询方式：①树形结构导航查询 (Navigate from tree top): 帮助用户从树状结构体系入手查询主题词的信息；②输词查询 (Search): 即直接在检索框中输词进行查询, 可以将欲查询的词限定在主题词、限定词或补充概念 (化学物质名称) 中某一种或全部, 还可以限定在《化学主题词表》的某一特定字段如化学物质登记号或酶编码 (CAS Registry/EC Number) 中进行查询。可选择三种匹配方式: Find Exact Term (精确词查询)、Find Terms with ALL fragments (查询词组中所有的词) 和 Find Terms with ANY Fragment (以词组中任意一个词入手进行查询) (图 2-3)。

利用 MeSH Browser 可获取完整的 MeSH 记录, 包括 MeSH 主题词数据 (MeSH Descriptor Data) 和 MeSH 树形结构 (MeSH Tree Structures), 其中前者主要包括词义范围注释 (Scope Note)、编目标引注释 (Annotation)、树状结构号 (Tree Number)、款目词 (Entry Term)、历史注释 (History note)、允许组配的副主题词 (Allowable Qualifiers) 等 (见表 2-4)。MeSH Browser 不直接和 MEDLINE 或其他数据库链接, 因此并不能代替 PubMed 系统。



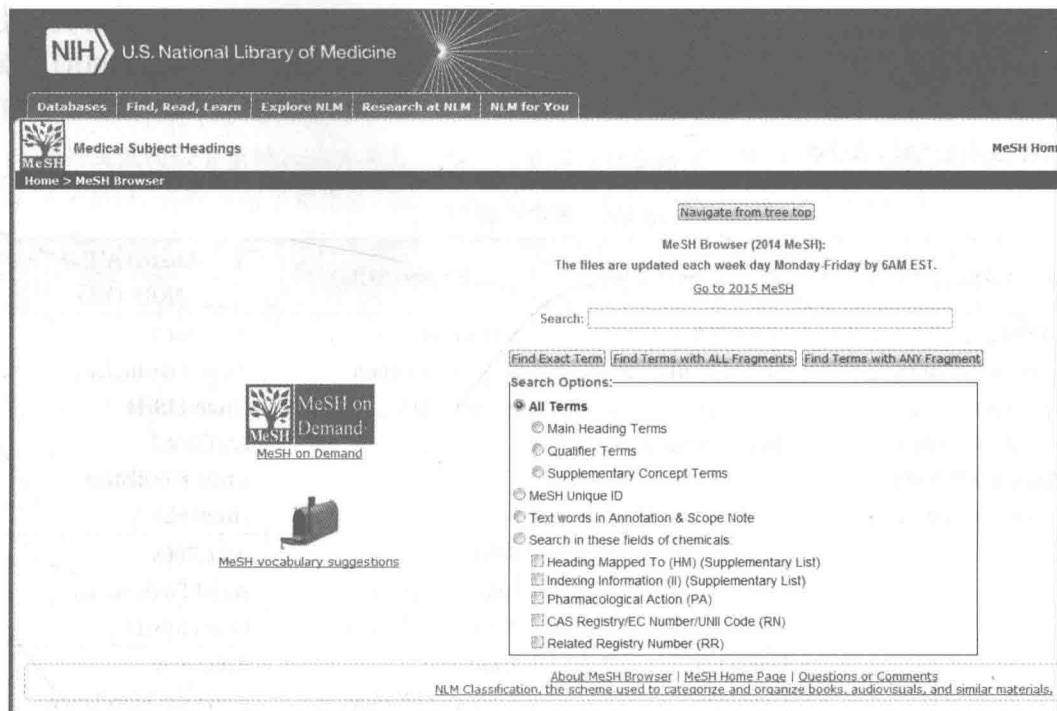


图 2-4 MeSH Browser 查询界面

四、一体化医学语言系统

一体化医学语言系统(Unified Medical Language System, UMLS)是美国国家医学图书馆自1986年开始进行的一项研究和开发计划,该计划旨在建立一个计算机化的可持续发展的生物医学检索语言集成系统和机读信息资源指南系统,使医疗卫生专业人员和研究人员能够通过多种交互检索程序,克服由于不同检索系统语言的差异性和不同数据库信息资源的分散性所带来的信息检索障碍,帮助用户从病案数据系统、书目数据库、事实数据库和专家系统等各种联机情报源中获取特定或综合性的信息资源。UMLS是一种全新模式的信息检索语言,它由超级叙词表、语义网络和专家词典三个部分组成,这三个部分是相互紧密联系和应用的整体。

1. 超级叙词表(Metathesaurus) 是一个生物医学概念信息数据库,包括生物医学概念的含义、属性、等级关系以及来源词表中概念之间的关系,在此基础上增加了概念的某些基本信息,并对不同来源词表的概念和术语之间建立同义关系及其他新关系。

超级叙词表来源于100多种词表和分类表,主要分为:叙词表如MeSH;疾病和病程分类表如《国际疾病分类》(ICD);专科或专业词表如《世界卫生组织药物副作用术语表》(WHOART);专家诊断系统如Dxplain;联机数据库如《孟德尔人类遗传学联机数据库》(OMIM);病案数据库术语表如《哥伦比亚长老会医学中心医学病种词典》(CPM)等。

超级叙词表是依据概念或含义(meaning)组织起来的,目的是将相同概念的交替名称和不同的词语形式联系在一起,识别不同概念之间的关系。UMLS规定将概念作为超级叙词表组织系统的中心。

表达同一概念可以有多种术语,而每个术语又有不同的词形变体。超级叙词表选择一个术语作为概念的名称,并把用于表达同一概念的来自各来源词表的所有其他词汇与这个优选的概念名称连接起来。在超级叙词表中,词汇变体和同义词是需要加以处理的两种主要的概念内关系。词汇变体包括拼写、词序、单复数变体以及缩写,主要出自来源词表。

对于同一概念的词间关系,超级叙词表采用4级表达模式,即概念(concept)(I级)、表达同

一概念的不同术语 (term) (Ⅱ级)、同一术语的不同变体即词串 (string) (Ⅲ级) 以及来自不同词表的原词 (atom) (Ⅳ级)。它们通过概念标识符 (CUI)、术语标识符 (LUI)、词串标识符 (SUI) 和原词标识符 (AUI) 相连, 将一个概念的多种不同术语、多个变异词串连同多个来源词表的原词有序地组织在一起 (表 2-5)。

表 2-5 概念的组织形式

Concept (CUI)	Terms (LUIs)	Strings (SUIs)	Atoms (AUIs) *RRF Only
C0004238 Atrial Fibrillation (preferred) Atrial Fibrillations Auricular Fibrillation Auricular Fibrillations	L0004238 Atrial Fibrillation (preferred) Atrial Fibrillations	S0016668 Atrial Fibrillation (preferred)	A0027665 Atrial Fibrillation (from MSH) A0027667 Atrial Fibrillation (from PSY)
		S0016669 (plural variant) Atrial Fibrillations	A0027668 Atrial Fibrillations (from MSH)
		S0016899 Auricular Fibrillation (preferred)	A0027930 Auricular Fibrillation (from PSY)
	L0004327 (synonym) Auricular Fibrillation Auricular Fibrillations	S0016900 (plural variant) Auricular Fibrillations	A0027932 Auricular Fibrillations (from MSH)

2. 语义网络 (Semantic Network) UMLS 的语义网络是建立语义类型及其相互关系的权威规则, 是为超级叙词表中的所有概念提供其语义分类以及概念之间可能存在的有用关系。在语义网络中, 语义类型可看作是节点, 节点与节点之间的关系即为语义关系, 语义类型和语义关系构成了网状的语义结构。

(1) 语义类型: 超级叙词表中的每一个概念至少被标引一种语义类型, UMLS 语义网络包括 133 种语义类型, 主要分为六大类: 生物体 (organisms)、解剖结构 (anatomical structures)、生物功能 (biological functions)、化学药品 (chemical)、客观物体 (physical object)、概念或观念 (idea)。

数据库中每一种语义类型记录通常包括 UI (唯一标识符)、STY (语义类型名称)、STN (语义类型的树状结构号)、DEF (语义类型的定义)、EX (超级叙词表中具有此种语义类型的概念的示例)、UN (语义类型用法注释)、HL (语义类型的等级关系链, 给予其父节点为 {isa}, 给予其子节点为 {inverse_isa}) 等。

(2) 语义关系: UMLS 语义网络包括 54 种语义关系, 其中最基本的语义关系是 “IS_A” 关系, 它表达语义类型之间的等级关系, 此外还有 5 大类非等级关系: 物理上相关 (physically related to)、空间上相关 (spatially related to)、时间上相关 (temporally related to)、功能上相关 (functionally related to)、概念上相关 (conceptually related to)。

数据库中语义关系记录通常包括 UI (关系的唯一标识符)、RL (关系名称)、ABR (关系缩写)、RIN (逆关系名称)、RTN (关系的树状结构号)、DEF (关系的定义)、HL (语义关系的等级链, 给予其父关系为 {isa}, 给予其子关系为 {inverse_isa})、STL (具有此种关系的语义类型) 和 STLB (不能继承该种关系的语义类型)。

3. 专家词典 (SPECIALIST Lexicon) UMLS 的专家词典始建于 1994 年, 包括大量生物医学词汇和常见英语词汇, 每个词典条目记录词汇的语法、形态和拼写信息, 为自然语言处理系统提供所需信息。

本章小结

文献信息检索语言是文献信息检索系统的标识系统,也是用户和文献信息检索系统进行交流的语言系统。本章简要介绍了文献信息检索语言的概念、意义、分类及应用等基本情况,详细介绍了中图法、DDC、NLMC等国内外著名的文献分类法,以及国际疾病分类法、国际系统医学术语集和医学主题词表。

(史继红)

思考题

1. 检索语言有何作用?
2. 按所描述的文献信息特征划分,检索语言有哪些类型?
3. 规范语言有何特点?
4. 主题词和关键词有何差别?
5. ICD有何作用?
6. SNOMED CT的顶级概念有哪些?
7. SNOMED CT可应用于哪些领域?
8. MeSH表有何作用?
9. MeSH表的概念体系由哪些类型的词组成?
10. MeSH Browser提供的查询方式?

第三章 图书馆信息资源利用

进入网络信息时代,图书馆的内涵、结构与功能也在悄然地演化,但是,她所承载的传承与传播知识的学术使命,依然是永恒的。因此,能否有效利用图书馆及其丰厚的资源,同样是这个时代医学工作者信息素养的体现。图书馆有选择地对文献资源和网络信息资源进行系统的、专业化的收集、加工、整理、贮藏或导航,并向读者提供各种文献信息服务,因此,要充分有效地利用图书馆的各种资源、服务。

第一节 馆藏资源的获取

一、馆藏目录查询

馆藏目录是查询图书馆文献收藏情况的工具。随着互联网应用的普及,基于 Web 的联机公共查询目录(Online Public Access Catalogue, OPAC)使读者可以通过互联网查询图书馆馆藏信息,了解图书馆所收藏的图书、期刊目录,被收藏的地点,以及当前的使用状态等,充分体现以用户为中心,使用更简单、方便,用户参与度更高,延伸了图书馆信息服务。

(一) OPAC 功能

1. 查询馆藏信息 通过系统提供的检索途径查寻图书馆图书和期刊的目录及馆藏的流通状态信息(如某本图书在馆、借出或被预约等状态)、已借出图书的应还日期、馆藏复本数、馆藏处理的状态信息(如订购中、在编处理中等)、期刊馆藏信息(如下一年是否有订购、最近到馆卷期、装订中的卷期、已经装订成册的卷期)等。

2. 查询读者辅助信息 注册读者可以查看自己借阅、续借、预约图书的记录信息,向图书馆推荐采购图书的记录等。

3. 流通功能 包括网上续借、预约和取消预约等。

4. 个性化信息服务 新一代 OPAC 系统一般都提供个性化的信息搜集、组织、推送服务,如 RSS (really simple syndication, 是在线共享内容的一种简易方式,也叫聚合内容)定制服务及时获取超期提醒、预约提醒及委托提醒等服务;个人虚拟书架可以将自己感兴趣或者关注的图书保存到收藏中,并可以对收藏进行分类管理。还可以对图书进行推荐,并查看图书的读者评价情况。

5. 其他功能 OPAC 系统还提供图书流通、书评、评价情况的统计查询,提供热门借阅、热门收藏以及其排行榜等。利用可视化技术提供借阅关系图等,有助于读者了解书与书之间的关系,了解阅读热点,更好地了解图书,方便读者使用图书。

(二) OPAC 检索方法

不同的 OPAC 系统提供的检索途径不完全相同,一般提供简单检索、多字段组合检索、全文检索、分类检索等途径。简单检索可以选择不同的检索字段对单个字段进行检索,如对书名字段检索,可以输入书名或部分书名进行检索。多字段组合检索提供多个字段,如书名、责任者、主题词、ISBN/ISSN、出版时间等的组合查询,字段之间是逻辑与的组配关系,一般用于精确查

找某一本书刊。全文检索支持布尔逻辑组配,提供任意字段、词汇的组合查询,可以进行逻辑与、或、非等组配方式筛选结果,使目录查询更精确、更灵活。有的 OPAC 系统还提供了分类浏览的功能,按照分类号逐级浏览,查找所属类目的馆藏资源,方便读者集中查找某一学科或者专题的文献。例如点击“R 医药、卫生”的子类目“R9 药学”,就可以查看图书馆收藏所有药学方面的图书。

二、资源发现系统

(一) 发现系统的定义

发现系统(discovery system)也称作统一资源发现系统,网络级发现服务等。学界对资源发现系统并没有明确的定义,归纳不同产品对资源发现系统功能的表述可以这样说:资源发现系统致力于从出版商、大学、公开的网站收集学术文献信息,形成中心知识库,通过预索引的方式为用户提供快速、简单、易用、有效的资源发现与传递服务。

发现系统是信息资源整合系统发展到一定阶段的产物。它以元数据集中索引为主要技术手段,对海量、异构的数字资源进行收集、聚合、索引,通过单一而强大的搜索引擎向用户提供统一检索和服务的系统。发现系统进行资源整合主要采用元数据集中索引的技术方式,通过抽取、映射和导入等手段对分布异构资源的元数据或对象数据进行收集和聚合,并将这些数据按映射转化规则转换为标准的格式纳入到元数据标准体系中,形成一个预聚合的元数据联合索引库,安装在本地系统或者中心系统平台提供统一的检索和服务。

(二) 发现系统的功能特色

发现系统可以将不同来源、不同类型的数据库聚合到一个平台内,包括图书馆自身的物理馆藏、数字馆藏、远程数据库、电子资源以及开放获取资源等,系统提供统一的检索界面。

发现系统的检索操作简单、方便、易用,一般在统一界面上提供一个类似于 Google 的简单检索功能的单一的检索接口,通过这一接口只需在一个检索框中输入相应的检索式,即可实现对不同数据库系统的一站式检索。

发现系统的检索结果可以按照不同的算法进行相关性排序。还可以对检索结果进行不同版本、不同类型的聚类显示,并以有序化的格式输出包括相关性排序和时间排序等。有些资源发现系统还对结果集提供可视化展示方式,便于发现资源间的联系。

(三) 常用资源发现系统

目前国内图书馆应用比较多的发现系统主要有,美国 Serials Solution 公司的 Summon 系统、以色列 Ex Libris 公司的 Primo 系统、美国 EBSCO 公司的 EBSCO Discovery Service (EDS) 系统、美国 OCLC 的 Worldcat Local 系统,美国 Innovative Interfaces 公司的 Encore 系统,以及中国超星公司开发的超星中文发现系统。各图书馆网站上发现系统的冠名各有不同,有的直接采用产品名称,有的则结合本校特色采用个性化的命名,如清华大学的“水木搜索”、北京大学的“未名学术搜索”等,其后台基本都是目前国际上主流发现系统之一。

三、移动图书馆

(一) 移动图书馆定义

移动图书馆源自英文“mobile library”,原指“汽车图书馆”,是为不能到达公共图书馆看书的民众设计的一种图书馆流动车。随着信息技术的发展和移动设备的普及,移动图书馆逐渐从实体的流动图书馆发展成为利用现代移动设备获取信息的新型服务方式。移动图书馆依托成熟的无线移动网络、互联网以及多媒体技术,使人们不受时间、地点和空间的限制,通过各种便携移动设备(手机、PDA、手持阅读器、平板电脑、MP4 等)方便灵活地利用图书馆的信息查询、浏览和获取资源内容等服务。

（二）移动图书馆服务模式

随着互联网技术的发展,移动图书馆服务模式也在不断改进,其服务模式主要包括短信服务、WAP 网站服务和 APP 服务。

1. 短信服务模式(short message service, SMS) 经过注册的用户可以享受图书馆的 SMS 服务。图书馆可以主动给读者发送新闻、讲座、预约到达、图书催还、过期罚款催缴等信息,而用户也可以按照一定的指令查看馆藏、借阅情况、续借、图书馆工作时间、参考咨询等信息。

2. WAP 网站服务 WAP 是终端无线通讯协议(Wireless Application Protocol)的简称,是在数字移动电话、因特网或个人数码代理(personal digital assistant, PDA)、计算机之间进行通讯的开放式全球标准。WAP 方式不受时间地点的限制,只要用户的终端设备上有 WAP 终端并安装好客户端软件,经过注册后,就能检索到图书馆的电子资源。图书馆建立的 WAP 网站除了为用户提供 OPAC 检索外,还提供电子书借阅、电子刊阅读、特色资源检索、虚拟参考咨询、在线服务、在线阅读、留言反馈、日志功能等服务,还可提供个性化推送、阅览室定位帮助和指南信息等高级服务。

3. 智能手机 APP 模式 APP 是 application 的简称,通常专指智能手机上的应用软件。这些智能手机除了可以通过传统的 WAP 方式上网查询外,还可以通过 Wi-Fi、3G 来访问网络。为此很多图书馆和机构都开发了专门的 APP 应用程序,供用户免费下载,向智能手机用户提供馆藏资源的检索功能,并向拥有有效证件的用户提供借阅信息查询、续借、预约等功能,用户也可通过 APP 查看图书馆的最新公告、讲座预告及各项服务帮助和指南。

4. 数据库的移动阅读和获取 随着移动互联网的发展,移动阅读市场的扩大,传统的出版商和数据库商也开始提供基于移动互联网的终端设备和数据库产品。如 EBSCOhost、PubMed 等数据库陆续推出了能够在移动设备上进行搜索、收藏甚至下载全文的数据库产品,如 EBSCOhost Mobile、PubMed Mobile 等可以直接通过智能手机检索并下载全文。

（三）超星移动图书馆

超星移动图书馆是以移动无线通讯网络为支撑,以图书馆集成管理系统平台和基于元数据的信息资源整合为基础,以适应移动终端一站式信息搜索应用为核心,以云共享服务为保障,通过手机、Pad 等手持移动终端设备,为图书馆用户提供搜索和阅读数字信息资源,自助查询和完成借阅业务,帮助用户建立随时随地获得全面信息服务的现代图书馆移动服务平台。超星移动图书馆主要有以下功能:

1. 馆藏目录检索 查询本馆纸质资源的基本情况,让读者方便快捷地实现对本馆资源的查询、预约、续借等操作。

2. 数字资源的检索与阅读 系统应用元数据整合技术对馆内外的中外文图书、期刊、报纸、学位论文、标准、专利等各类文献进行了全面整合,在移动终端上实现了资源的一站式搜索、导航和全文获取服务。目前,实现与 700 多家图书馆馆藏书目系统、电子书系统、中文期刊、外文期刊、外文数据库系统集成,读者直接通过网上提交文献传递申请,并且可以实时查询申请处理情况,以在线文献传递方式通过所在成员馆获取文献传递网成员单位图书馆丰富的电子文献资源。考虑到手机阅读的特点,超星移动图书馆专门提供 3 万多本 e-pub 电子书和 7800 多万篇报纸全文供手机用户阅读使用。

3. 个性化定制服务 超星移动图书馆集成 RSS 订阅功能,为用户提供个性化信息服务。包括电子书籍、报纸、杂志、视频、资讯等近 30 种频道分类,用户可以有针对性地阅读自己所需的信息,提供多来源信息的个性化阅读体验。



第二节 图书馆信息服务

一、文献借阅

文献借阅是图书馆最基本的服务功能。很多大学图书馆将藏书库与阅览室合二为一,采取开架管理,读者可以进入书库直接查找和阅览图书,借还书需要凭证到流通台办理手续。为保证流通,使更多的读者都能利用馆藏文献,读者需要熟知自己的借阅资格(借阅数量、期限等)及其他限制。为保证读者能够利用到馆藏文献,许多图书馆对某些文献(如重要的专业著作、新出版的图书)保留一本作为馆藏本,和辞典、百科全书、药典之类的工具书一样不外借。还有些图书馆收藏古籍、学位论文、会议资料和其他特种文献,有限制的提供阅览服务。由于期刊每卷每期汇集了多篇论文,具有利用率高、时效性强、复本数少等特点,图书馆一般都不外借,只提供馆内阅览和复印服务。

二、参考咨询与读者培训

参考咨询是读者在利用图书馆和电子资源的过程中遇到各种疑难问题时,由图书馆员利用各种参考工具、检索工具,为读者解答和解决问题的一种服务方式。常见的咨询方式有:馆内现场咨询、电话咨询、邮件咨询、实时在线咨询、图书馆主页的留言簿和常见问题解答(FAQ)等形式。近年来随着互联网的应用与普及,虚拟参考咨询得到了广泛的应用,并取得非常好的服务效果。虚拟参考咨询是图书馆提供的以网络为依托,以本馆馆藏和网上数字化信息资源为基础,通过邮件、实时交流等为读者提供的不受时空限制的参考咨询服务。读者可以通过网络提出咨询问题,请求在线馆员给予解答。

读者培训是图书馆有计划、有目的地开展的,旨在提高用户的信息意识和检索技能,使其能充分利用图书馆及其信息资源的教育活动。大学图书馆主要有以介绍图书馆利用基本知识为主的新生入馆教育培训和以推广宣传利用某些资源或讲座为主的一系列培训活动。

三、科技查新

科技查新简称查新,是指查新机构根据查新委托人提供的有关科研资料,通过系统全面的文献检索,查证其课题、研究内容或科研成果,是否具有新颖性,并出具相关佐证文献资料的文献调研工作。查新是文献检索和情报调研相结合的情报研究工作,它以文献为基础,以文献检索和情报调研为手段,以检出结果为依据,并与课题查新点对比,对其新颖性作出结论并出具查新报告。通过查新能为科研立项,科技成果的鉴定、评估、验收、奖励,专利申请等提供客观依据;也能为科技人员进行研究开发提供快捷、可靠、丰富的信息。国内一些大学图书馆获得了有关机构的认定,成为有资质的查新机构,为读者提供科技查新检索服务。

四、馆际互借与文献传递

馆际互借是图书馆之间相互利用对方馆藏来满足本馆读者需求的一种资源共享服务。馆际互借一般针对图书,是一种返还式文献资源共享方式,是图书馆根据读者需求,将本馆没有收藏的图书,从其他收藏馆借阅过来提供读者使用的一种服务。文献传递是在馆际互借基础上发展起来的,是馆际互借的一种,是非返还式的文献资源共享方式。读者申请文献传递时需要提供所需文献的篇名、作者、刊名、卷、期、起止页码等完整的题录信息。图书馆将查获到的文献原文根据读者要求通过传真、邮寄、电子邮件、网络工具等传递方式送达。

目前高校系统图书馆、国家图书馆、国家科技图书文献中心(NSTL)、中国高等教育文献保

障系统(CALIS)等机构相继开展了馆际互借与文献传递服务,推出文献传递服务平台,数字图书馆平台也推出了文献传递功能,可以在检索之后直接在线提出文献传递申请,系统自动将读者所需文献发送到读者电子邮箱。

五、学科服务

学科服务是图书馆为适应新的信息环境,以用户的需求为中心而推出的贴近用户一线的新的服务模式,它打破了传统的按照文献的工作流程组织科技信息的方式,而是按照科学研究的学科、专业、项目、课题等来获取、组织、检索、存储、传递与利用信息资源,从而使信息服务学科化、服务内容知识化。

目前,大学图书馆学科服务的主要内容有:对用户开展电子资源利用情况的调查,征求对口院系教师对文献资源的订购意见,联系院系与图书馆合作订购数据库,编写数据库使用指南,向用户发送信息通报,上门进行读者培训和资源介绍,咨询解答,建设电子资源导航,开辟网上学科博客,为院系用户提供专题文献信息、参与院系教授课题组的情报服务、知识服务和决策支持等方面的服务。

学科服务由图书馆学科馆员来推动,通过电话、邮件、学科博客、研究室、课题组等方式,将图书馆信息服务延伸到用户之中。充分利用图书馆学科服务可以获得更具有针对性、个性化的服务。

六、信息共享空间

信息共享空间(information commons, IC)有的也称作学习共享空间(learning commons, LC),是一种经过特别设计的一站式服务中心和协同学习环境,是大学图书馆创新服务模式。信息共享空间综合了图书馆空间、因特网、计算机软硬件设施和图书馆知识库等资源(包括馆藏纸本、数字化和多媒体等各种信息资源),并在技能熟练的图书馆参考咨询馆员、计算机专家、多媒体工作者和指导教师的共同支持下,为读者的学习、讨论和研究等活动提供一站式服务,目的在于培育读者的信息素养,促进读者学习、交流、协作和研究。

第三节 信息资源共享

一、CALIS资源与服务

CALIS(<http://www.calis.edu.cn/>)的简称是中国高等教育文献保障系统(China Academic Library & Information System, CALIS),是经国务院批准的我国高等教育“211工程”、“九五”、“十五”总体规划中三个公共服务体系之一。CALIS的宗旨是把国家的投资、现代的图书馆理念、先进的技术手段、高校丰富的文献资源和人力资源整合起来,建设以中国高等教育数字图书馆为核心的教育文献联合保障体系,实现信息资源共建、共知、共享,以发挥最大的社会效益和经济效益,为中国的高等教育服务。

通过CALIS文献信息服务网络平台可以获取全国高校图书馆、国家科技图书文献中心、国家图书馆的相关文献(期刊论文、会议论文、学位论文、图书章节等)。CALIS服务平台按照服务内容可分为咨询服务和检索服务。咨询服务包括:期刊导航、数据库导航、图书馆导航、e得文献获取、联合问答(e问)、科技查新、收录引证、课题咨询等;检索服务包括:学术搜索引擎(e读)、书刊联合目录、外文期刊网、中外文学学位论文、电子教参书籍、高校特藏资源、百万电子图书等。

1. e读学术搜索引擎服务 整合全国800多家高校纸本资源和电子资源,揭示资源收藏与



服务情况,通过一站式检索从海量资源中快速发现与获取有用的信息。e 读整合了联合目录、外文期刊网、学位论文、教学参考书、特色库、古籍等各类资源合计 9000 余万条数据,可集成馆藏 OPAC、电子资源全文阅读、章节试读、无缝链接 CALIS 馆际互借体系。

2. e 得文献获取服务 e 得(易得)是为读者提供“一个账号、全国获取”、“可查可得、一查即得”一站式服务的原文文献获取门户。e 得门户集成了电子原文下载、文献传递、馆际借书、单篇订购、电子书租借等多种原文获取服务。结合专业馆员提供的代查代检服务,可在 CALIS 各类检索工具覆盖的文献资源之外,帮助读者在全国、乃至全世界范围查找并索取包含中外文的图书、期刊、学位论文、会议论文、专利标准等各种类型的电子或纸本资源全文。支撑 e 得全文服务的不仅有 800 多家 CALIS 高校成员馆,还有以国家图书馆、上海图书馆为代表的众多公共图书馆, NSTL、中国科学院图书馆为代表的各类学术图书馆或文献服务机构, CASHL、教育部外国教材中心、CADAL 等为代表的教育部资源共享项目,以及以方正阿帕比、同方知网、维普资讯、万方数据等为代表的国内资源数据库商。

3. 外文期刊网服务 CALIS 外文期刊网是全面揭示国内高校外文期刊的综合服务平台,是获取外文期刊论文的重要途径。外文期刊网包含 10 万多种纸本期刊和电子期刊,7000 多万篇文章检索信息,100 多个全文数据库、11 个文摘库的链接,196 个图书馆的馆藏纸本期刊信息,497 个图书馆购买的电子期刊信息。全面揭示高校纸本期刊和电子期刊,为用户提供一站式期刊和文章检索及全文链接服务。

二、NSTL 资源与服务

国家科技图书文献中心(National Science and Technology Library, NSTL, <http://www.nstl.gov.cn>),是一个虚拟的科技文献信息服务机构,成员单位包括中国科学院文献情报中心、工程技术图书馆、中国农业科学院图书馆、中国医学科学院图书馆等。该中心根据国家科技发展需要,按照“统一采购、规范加工、联合上网、资源共享”的原则,采集、收藏和开发利用理、工、农、医各学科领域的科技文献资源,面向全国开展科技文献信息服务。NSTL 提供的服务主要包括:文献检索、重点领域信息门户、期刊浏览、全文文献获取、引文检索、代查代借、参考咨询、热点门户、预印本服务等。

文献服务是 NSTL 的一个主要服务项目。具体内容包括:文献检索、全文提供、网络版全文服务、目次浏览、目录查询等。非注册用户可以免费获得除全文提供以外的各项服务,注册用户同时可以获得全文提供服务。

第四节 学术期刊与参考工具书

一、学术期刊

1. 学术期刊概述 期刊是指有固定名称、版式和连续的编号,每一期可刊载由不同著者撰写的多篇文章或论文,定期或不定期长期出版的连续性出版物。学术期刊以学术交流为目的,主要报道科研方面的学术研究成果,是科学家之间最为重要的学术交流媒介。

学术期刊按内容和性质划分有信息量大、价值高的学术性期刊(学报、通报、汇刊、评论、进展等),报道消息的快报性期刊(通讯(letters)、短讯(news)等),获取文献线索的检索性期刊(文摘(abstracts)、索引(index)),保存历史记录的资料性期刊(data、event)等。

按照学术期刊的编辑出版机构,主要有学会、协会等学术团体出版的期刊,商业出版社出版的期刊,政府机关出版的期刊,大学学报等。识别期刊的主要依据有:名称,出版的年、卷、期,国际标准刊号(ISSN)等。国外把丛书(series)也看作连续性出版物,常使用 ISSN 号。

2. 学术期刊的评价 期刊的特点是内容新颖、信息量大、出版周期短、传递信息快、传播面广、时效性强,能及时反映国内外各学科领域的发展动态。期刊出版种数浩大,每个学科专业都有自己的核心期刊。所谓专业核心期刊是指刊载该专业论文数量较大,学术水平较高,能反映本学科最新研究成果及本学科前沿研究状况与发展趋势的,备受该学科专业读者重视,影响因子系数高的学术期刊。核心期刊与非核心期刊的划分是相对的、动态变化的。国内外认定核心期刊的工具较多,各单位有不同的认定规定,常用的期刊测评方法有:

(1) 载文量法:将某一领域的期刊按相关载文量的多少递减排列,然后累计排在前面的 n 种期刊的载文量。当此期刊载文量与所统计的全部期刊总载文量达到了选定的要求,例如80%,则前 n 种期刊为核心期刊。

(2) 文摘量法:根据被书目文献数据库摘录情况将期刊依次排序,凡期刊中被摘录和索引的论文数量较大者,可选为核心期刊。

(3) 流通量法:对馆藏期刊在一定时间内的外借次数、阅览次数、复制次数进行统计分析,流通量高的即为核心期刊。

(4) 引文量法:根据期刊上文献被引用情况的统计,如按文献被引量、影响因子、即时被引指数等将期刊排序,靠前的、被引用率高的被认为是某一学科的核心期刊。

(5) 专家评审法:由各学科领域的专家、主编等对各学科期刊打分进行评选。

(6) 综合法:由于以上各种方法都有各自的局限,不能满足对期刊评价的要求,因此将上述多种评价指标科学地组织在一个评价体系中,即综合评价法。

国内最常用的期刊评价工具有北京大学编制的《中文核心期刊要目总览》(每四年修订1次)、中国科学技术信息研究所的《中国科技期刊引证报告(核心版)》的统计源期刊目录、《中国科学计量指标:论文与引文统计》等;国际上的科技期刊则通用美国科学情报研究所(ISI)的期刊引用报告(Journal Citation Reports, JCR)。

二、参考工具书

1. 参考工具书概述 参考工具书是汇集某一学科范围的知识信息,按照一定的体例和检索方法编排而成的专供查考资料或资料线索的特种图书。参考工具书与检索工具书都是信息检索的必备工具,都具有检索功能,但是二者有一定的区别:检索工具书是用以报道、存储和检索各种文献的二次文献,提供文献资料的线索与出处,在检索途径的设计上更系统、更便捷;而参考工具书是三次文献,提供使用者确切的数据、事实或结论,表述确切的知识,在内容的表述上更注重科学性。与普通图书相比,参考工具书具有查考性、概括性和易检性的特点。

2. 参考工具书类型 参考工具书的种类繁多,依据选材、加工、编排方式不同,解答疑难的范围、角度和要求不同,形成很多不同的类型。

(1) 词典:又称辞典,是解释词语的概念、意义和用法的工具书。词典的类型很多,从文种上划分有单语、双语和多语对照词典;从注释繁简、篇幅长短上划分有词典、词汇;从取材范围、编辑目的上划分有综合性、专业性、缩略语等。

(2) 百科全书:是百科知识的总汇,是一种大型的综合性工具书。百科全书多采用条目形式来解释各学科知识的定义、原理、方法、历史和现状等,同时具有完善的检索体系,读者能够迅速准确地查获答案。就其收录范围、学科性质和通俗程度,具有综合性、专业性和通俗性三种。

(3) 年鉴:是一种每年出版一次的连续出版的参考工具书。它记录某一领域一年内的科学进展、学术活动、出版动态、统计数据和大事件的事实信息。年鉴编辑单位具有一定的权威性,多为政府有关部门、学术团体或研究机构。年鉴出版及时,报道新颖,以记事为主,不作更多的叙述和解释,具有连续性和时效性特点。年鉴就其编辑内容、性质和用途可分为记事、综述和



统计年鉴,就其取材范围可分为综合性、专门性和区域性年鉴。

(4) 手册:手册的名称很多,有指南、便览、要览、一览、宝鉴、必备、大全、全书等。不管名称如何,它们都是汇集某一方面的基本知识和基本数据,专供随时查阅的参考工具书。手册一般内容实用,文字简练,图表简明,编排体例简便、易查。

(5) 名录:包括人名录、机构名录。是查考人物、机构名称,了解其基本情况的参考工具书。人名录(Who's Who)又叫“名人录”,源于名人传记,是一定范围内知名人物简传的汇集。机构名录是各种组织机构名称及其基本信息的汇编。

(6) 图谱:是汇集某一学科的事物,用图像形式绘录或摄制下来,加以分类编排的一种直观性的参考工具书。医学图谱用图像配以文字来表现医学实体、表达操作规程或反映疾病的地域分布情况。医学图谱品种多样,有解剖图谱、病理组织学图谱、细胞图谱、X线图谱、超声图谱、手术图谱等。网络图谱与印刷本图谱相比具有容量大、可检索、可交互操作等特点,图像更加形象直观,并且能够动态立体显示,对于教学和研究帮助极大。

(7) 药典:是一个国家记载药品规格、标准的法典。大多数由国家组织的药典委员会编制,并由政府颁布,具有法律约束力。药典记载那些疗效确切、副作用低、质量稳定的药物及其制剂,规定其质量标准、杂质鉴定、含量测定,并作为药物生产、检验、经营、使用、监督的依据。

3. 医学参考工具书举要 进入移动互联网时代,除了传统的印刷本参考工具书,网络参考工具书,也出现了通过智能终端进行移动阅读的参考工具书。数字参考工具书存储容量大、查询速度快、检索途径多、内容更新快,并具有多媒体效果,大大方便了读者的查询与利用。

(1) 《道兰图解医学词典》(*Dorland's Illustrated Medical Dictionary*):是一部享誉世界的权威性医学词典。初版于1900年,2013年出版了第32版手机版。该词典收词广泛、释义精确,补充新词及时,编有医学词源,并广泛使用了互见参照,附有丰富的插图。

(2) 《中国医学百科全书》:全书由卫生部组织4000多名专家编写的一部具有权威性的大型医学参考工具书,1983年起陆续由上海科技出版社出版,历经十余载,全书93个分卷陆续问世。全书共收录词条2万余条,条目编排以知识体系为主,文字简明扼要,内容新颖可靠,对广大医药卫生人员具有较高的参考价值。

(3) 《医学百科全书》(*Medical Encyclopedia*, http://www.nlm.nih.gov/medline_plus/encyclopedia.html):由美国国家医学图书馆(NLM)的MedlinePlus提供的免费使用的医学百科全书,有4000多条涉及疾病、试验、症状、损伤和外科等方面的条目,还有大量的照片和插图。

(4) 《默克诊疗手册》(*The Merck Manual of Diagnosis and Therapy*, <http://www.merckbooks.com/manual/index.html>):1899年初版,定期修订,目前出版到第19版。是连续出版最久、在全世界使用最广泛的医学手册。该书几乎囊括了人类所患的各科疾病,提供简明、完整、准确、有用的临床信息。

(5) 医学年鉴:我国每年出版的医学年鉴有20多种,有综合性的也有专题性的,如《中国医学科学年鉴》《中国卫生年鉴》《中国药学年鉴》《中国外科年鉴》等。

(6) DIRLINE (*Directory of Information Resources Online*, <http://dirline.nlm.nih.gov>):由NLM编制的查询美国和其他国家卫生与生物医学机构的名录。

(7) 公共卫生影像库(Public Health Image Library, PHIL, <http://phil.cdc.gov/Phil/home.asp>):是由美国疾病预防控制中心(CDC)主办的,服务对象为教育工作者、生物医学专业科研人员、医学生等。该网站搜集了大量与医学有关的包括照片、演示图例、多媒体文件等影像资料。互联网上的医学影像资源多分布在各个学科分类,变化情况比较大,小型分支学科的医学影像库较多,公共卫生影像库属于其中的综合性大型医学影像库。

(8) 《中华人民共和国药典》:2010年版为新中国成立以来第9版药典。本版药典分一部、二部和三部,收载品种总计4567种,其中新增1386种。药典一部收载药材和饮片、植物油脂和

提取物、成方制剂和单味制剂等,品种共计 2165 种,其中新增 1019 种(包括 439 个饮片标准)、修订 634 种;药典二部收载化学药品、抗生素、生化药品、放射性药品以及药用辅料等,品种共计 2271 种,其中新增 330 种、修订 1500 种;药典三部收载生物制品,品种共计 131 种,其中新增 37 种、修订 94 种。

第五节 医学电子书

一、电子书概述

电子书(electronic book)也称数字图书或 ebook,是指以数字代码方式将图、文、声、像等信息存储在磁、光、电介质上,通过计算机或类似设备使用,并可复制发行的大众传播体。

(一) 电子书的特点

1. 海量存储、节省资源 以数字化的形式将信息高密度地储存在计算机信息存储介质上,不需要纸张印刷,保护森林资源,是一种环保的绿色产品,也不占空间。

2. 功能齐全、使用方便 电子书集文字、声音、图像等多媒体信息于一体。能够最大限度地调动读者的感官,电子书的信息资源便于复制、检索,统计功能强大,方便读者的使用。

3. 快速传播、便捷流通 电子书不需要印刷,发行周期短,更新速度快,可以快速地传播和便捷地流通,不受时间和空间的限制。并且方便携带,无论是基于网络使用、还是使用移动存储或是便携式阅读器,都具有携带方便的特点。

4. 价廉物美、资源丰富 在互联网上有丰富的免费电子书资源。而收费的电子书其价格也比传统的图书便宜,一般只是印刷型图书价格的 20% 左右。此外,印刷型图书一次只能供一位读者使用,而大多数电子书可供多位读者同时使用,性价比高。

(二) 电子书的类型

1. 按照载体形式划分 按照载体形式电子书可分为基于网络的电子书和离线型电子书。

基于网络的电子书存放于网络服务器,使用通用浏览器或专门阅读软件阅读,需要通过网络到提供电子书的服务器上下载,借助安装在计算机上的电子书软件进行阅读。部分网站的电子书也可下载到终端设备中进行离线使用。

离线电子书是指利用移动阅读设备下载或借阅后、不依赖于网络就可以阅读的电子书。离线电子书必须基于移动阅读终端设备而使用,常见的设备包括专用的电子书阅读器、手机、平板电脑和 U 盘等。电子书阅读器是一种便携式、配合电子书离线阅读的数码产品,可以存储一定数量的电子书,随时随地进行阅读,国内外均有专门厂商的产品推出,例如亚马逊的 kindle 电子书阅读器、汉王科技的汉王电纸书等。

2. 按照出版来源划分 按照出版来源电子书可以划分成两大类:一类是出版社自行制作出版的电子书数据库,同时出版印本图书和电子书,拥有图书的印本和电子版的双重版权,独立发行,自行提供检索服务平台,其电子书集成在出版社的网络平台上,能同时检索该出版社出版的丛书、专著、参考工具书等不同类型的图书,并且文中参考文献可相互链接,如英国的牛津大学出版社、德国的 Springer 公司等公司出版的电子书。一类是数据集成商将多个出版社的电子书整合在同一个数据平台上为用户提供服务,这类电子书品种多、数量大,覆盖多个学科领域,但由于所收录的电子书并不专属于集成商,因此一本图书可被多个同类数据库所收录。这类数据库涉及作者和出版社向集成商授权图书电子版权的问题,因此,有些数据库存在并发用户或打印、下载图书量的限制。如超星公司的读秀学术搜索就属于此类电子书。



二、电子图书举例

(一) 读秀学术搜索

读秀学术搜索(<http://www.duxiu.com/>)由北京世纪超星信息技术发展有限责任公司创建,由全文数据及元数据组成的超大型数据库,由海量全文数据及资源的基本信息组成,以330万种中文图书、10亿页全文资料、6700多万种期刊元数据为基础,为用户提供深入内容的章节和全文检索,部分文献的原文试读,将图书馆纸质图书、电子图书、期刊、报纸、学位论文、会议论文等各种学术资源整合于同一数据库中,统一检索,使读者在一个检索平台上获取所有学术信息。

读秀图书频道为读者提供330万种中文图书的查找,同时可以在同一平台检索纸本书、电子图书。读者可以了解图书收藏情况,在图书馆拥有的纸本图书可以通过OPAC借阅,馆藏电子图书可以在线下载全文;图书馆没有的电子图书可以进行文献传递、读者互助获得部分篇章,也可以查看其他图书馆馆藏、推荐图书馆购买、按需印刷或者链接到网络书店等。

利用读秀图书检索有快速检索、高级检索、专业检索、分类导航等四种检索方式。其中快速检索是读秀图书检索的默认方式,简单、快速直接;高级检索提供书名、作者、出版社、主题词、年代等详细信息,更加精准定位图书;专业检索使用逻辑算构建检索表达式,适合专业人士;分类导航适合读者检索目标不是很明确,浏览某主题分类的图书。

(二) Thieme E-Book Library

Thieme E-Book Library(<http://ebooks.thieme.com/bookshelf>)是一个内容持续增加的在线参考书资料库,目前拥有Color Atlases和Flexibook系列的彩图教科书,所收录基础科学和医学内容,共计四大类:解剖学、放射学、临床医学、基础科学。其Color Atlas系列图书学科涵盖医学院校开设的所有临床医学和科学课程,用简洁的内容表达复杂的学科内容,方便学生学习理解和记忆,书籍内附精美图片、文字介绍清楚,阐明主题,不但是图文并茂而且图文相互对照。提供字顺和分类浏览、简单检索、短语匹配检索等。检索结果提供在线阅览、下载离线阅读。下载离线阅读需要利用其iPublishCentral Reader阅读器进行阅读。

(三) McGraw-Hill 医学在线电子图书

McGraw-Hill是一个教育、信息及金融服务机构。其McGraw-Hill Medical(<http://mhmedical.com/umbrella-index.aspx>)是其出版的在线医学资源。它是一个为医学生、住院医师、临床医师、研究人员以及所有医学相关专业工作人员提供众多医学信息的网络平台,其中内容包括知名医学著作、医药信息、及时更新的医学资讯、数以万计的图片及图表、互动式课程定制和学习、自我测试和评估、病例、诊断工具、全面的搜索功能。

其Access电子图书产品包括AccessMedicine、AccessSurgery、AccessAnesthesiology、AccessEmergency、AccessPediatric等数据库。其中AccessMedicine包括医学图书、多媒体资源、药物库、健康时讯、实验室诊断与疾病诊断、实践指南、精要概览、病例分析、病患教育等内容。其医学图书包含了40多本由麦格劳-希尔公司出版的经典医学教科书,包括Harrison's Online(其中包含《哈里森内科学》)、CURRENT Medical Diagnosis & Treatment等,及非常著名的LANGE Basic Science和LANGE CURRENT Clinical Science系列。其Reading界面罗列出Accessmedicine的所有图书。检索方式分为简单检索和高级检索。

(四) 牛津医学在线

牛津医学在线平台(<http://oxfordmedicine.com/>)由牛津大学出版社推出,平台囊括了牛津大学出版社的500多种高质量的医学书籍,这些医学书籍提供网络浏览,并且其内容不断更新。牛津医学在线电子图书按类型主要包括Oxford Medicine Textbooks(牛津医学教科书系列),Oxford Medical & Nursing Handbooks Online(在线牛津医学手册),The European Society of

Cardiology Textbooks(欧洲心脏病学会 ESC 在线教科书系列),美国医学会医学论文写作指南等类型在线图书。其中,以牛津医学教科书系列图书最为常用,目前出版了第5版,内容涵盖内科临床实践及其亚专科的全部内容,并且随时更新。

牛津医学在线的电子图书主要提供按学科主题浏览、按职业阶段浏览、按电子书系列浏览三种浏览方式,除此之外,牛津医学在线还在页面提供快速检索和高级检索的链接。牛津医学在线的电子图书不支持整本书的下载,对于检索结果,可以对其进行打印、保存、Email 或添加书签等操作,还可以点击 Download chapter(pdf)链接下载章节的 PDF 全文,以方便离线浏览。

(五) Ovid 医学电子图书

Ovid 医学电子图书是 Ovid 平台上收录的 LWW 出版社的内科、外科、肿瘤、妇产科等各类医学电子图书,数量超过千种,很多图书经 Doody 评级,并获得很高的星级。Doody Enterprises, Inc. 公司主要针对医学、护理以及有关的保健学书籍提供国际上领先的权威和专业质量评估。其 Doody's Star Rating. 与来自问卷调查中的计分相关联,是由评论者在评定书目的过程中完成的(5 星级是最高的评定级别)。评分体系能确保每个评论尽可能地公正客观。

在 Ovid 检索平台点击 Books@Ovid,即可进入 Ovid 医学电子图书数据库。

(六) 其他医学电子图书

除了前面介绍过的电子图书数据库外, Springer、Elsevier、Karger 等著名出版社在各自的电子出版物平台上提供了大量的电子图书。

1. Springer 电子图书 Springer 电子图书数据库中包含超过 56 000 种电子图书、电子丛书系列和电子参考工具书,并每年递增多达 7 000 种,包括 12 个英文学科和 5 个德文学科电子图书。Springer 的电子图书数据库除了提供与 Springer 纸版图书相同的高质量制作和插图外,同时还与 Springer 其他数据资源整合在 Springer Link 平台上,具有强大的在线检索功能和深度链接功能。Springer 的电子图书数据库可进行全文检索功能,通过整合平台,可同时检索所有 Springer 出版的内容。

2. Karger 电子图书 Karger(<http://www.karger.com>)是瑞士一家著名生物医学出版公司,其出版品以医学专业为主,是目前全球为数不多完全专注于医学专业领域的出版社。Karger 电子图书的学科领域涵盖了医学的传统领域及现代医学的最新发展,从肿瘤学、内分泌学、肾脏学、细胞生物学、神经系统科、血液学到遗传学,读者都可以从 Karger 的书籍或者期刊中得到所有生物医学领域最新的发展、应用及研究信息。Karger 提供各年度所有最新出版的图书丛书超过 41 个丛书系列和 150 多本非丛书系列。大部分的 Karger 丛书也都被收录于 MEDLINE 数据库里。

3. ClinicalKey 电子图书 ClinicalKey 电子书是 ClinicalKey 平台上的电子图书部分,包含 Elsevier 经典医学图书超过 1 100,其中不乏大量的医学圣经典图书,如 *Cecil Medicine*(《西氏内科学》)、*Nelson Textbook of Pediatrics*(《尼尔森儿科学》)、*Miller's Anesthesia*(《米勒麻醉学》)等,还有 130 余种医学教材,如 *Gray's Anatomy*(《格氏解剖学》)等。

本章小结

本章介绍了图书馆馆藏资源的获取方式包括 OPAC 检索、资源发现系统和移动图书馆系统的概念、功能,介绍了图书馆提供的各种类型的信息服务如文献借阅,参考咨询与培训,馆际互借和文献传递,科技查新、信息共享空间等以及更广泛的信息资源共享,介绍了学术期刊及参考工具书以及电子图书的概念、类型和重要的医学电子图书。



思考题

1. 什么是 OPAC? 它具有哪些功能?
2. 什么是资源发现系统? 它具有什么特点?
3. 移动图书馆服务模式都有哪些?
4. 图书馆都能提供哪些信息服务功能?
5. 利用读秀学术搜索查找人民卫生出版社出版的《乳腺疾病病理彩色图谱》一书。
6. CALIS 平台上有哪些服务功能?
7. 什么是学术期刊? 期刊的评价方法有哪些?
8. 参考工具书具有什么特点? 常用的医学参考工具书有哪些类型?
9. 电子图书的类型有哪些? 简述常用的医学电子图书。
10. 简述超星移动图书馆的主要功能。

第四章 中文数据库检索

中文信息资源内容丰富、种类繁多。据国家新闻出版总署统计,2013 年我国出版自然科学、技术类科技期刊 4944 种、医药卫生类图书 1.8 万余种。中文信息资源没有语言障碍,是学习、科研、临床实践中经常使用的信息资源。常用的中文文摘和全文数据库主要包括中国生物医学文献服务系统、中国医院知识总库、万方医学网、维普期刊资源整合服务平台等,本章分别予以介绍。

第一节 中国生物医学文献服务系统

一、概 述

中国生物医学文献服务系统(简称: SinoMed)由中国医学科学院医学信息研究所开发研制。在中国生物学文献数据库的基础上,收录内容和检索功能不断扩展,增加了北京协和医学院博硕士学位论文库、中国医学科普文献数据库以及西文、日文、俄文期刊文献,是集检索、免费获取、个性化定题服务、全文传递服务于一体的生物医学中外文整合文献服务系统。可通过镜像站点、网络版包库等方式使用。中心网站为: <http://www.sinomed.ac.cn>。

SinoMed 收录预防医学、基础医学、临床医学、药学等生物医学各个领域的文献,是综合性生物医学文献服务系统。其中, SinoMed 收录 1978 年以来 1800 余种中国生物医学期刊以及汇编、会议论文文献题录 820 余万篇, 1989 年以来中文参考文献 400 余万篇,并不断更新和增加;中国医学科普文献数据库收录 2000 年以来国内出版的医学科普期刊近百种;北京协和医学院博硕士学位论文库收录 1981 年以来协和医学院培养的博士、硕士研究生学位论文,可在线浏览全文。此外, SinoMed 系统还收录了西文期刊、英文文集汇编、英文会议文献以及日、俄文期刊等文献,目前共可检索 8 个中外文数据库,具体见图 4-1。

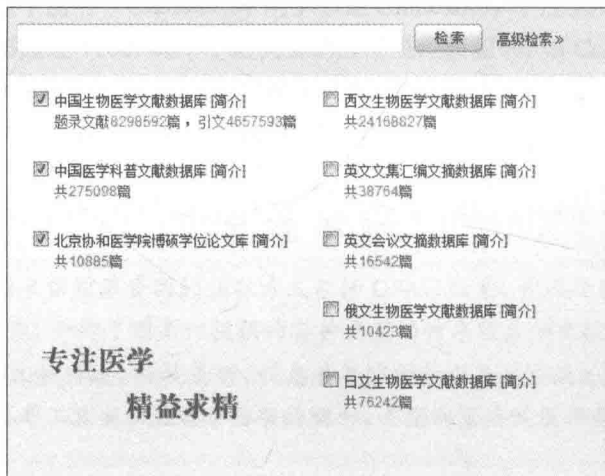


图 4-1 SinoMed 系列数据库资源

SinoMed 与 PubMed 检索系统具有良好的兼容性, SinoMed 全部题录均严格依据美国国立医学图书馆(NLM)的医学主题词标引规则, 采用《中文医学主题词表》(CMeSH, NLM 编制的《医学主题词表(MeSH)》中译本)、中国中医科学院中医药信息研究所的《中国中医药学主题词表》进行主题标引。同时, 采用《中国图书馆分类法·医学专业分类表》进行分类标引。

除题目、摘要、刊名、作者、作者单位、基金、参考文献等常规检索入口外, 提供特色的智能检索、主题词表/分类表辅助检索、第一作者检索、第一机构检索、限定检索、定题检索、多知识点链接检索、检出结果聚类分析等功能, 支持截词检索和各种逻辑组配检索, 具有多种排序、显示、输出功能, 具有检索策略的修改、保存、调用等功能。目前, SinoMed 题录数据与维普科技期刊全文数据库无缝链接, 1989 年以来的全文可直接通过链接获取全文。同时, SinoMed 提供西文期刊免费全文直接获取、非免费全文多渠道链接及在线索取等服务, 方便用户经济、便捷地获取全文。

二、检索方法和技巧

SinoMed 系统平台包括 8 个中外文数据库, 各子数据库的检索功能基本相同, 以中国生物医学文献数据库为例, 主要包括以下检索方式。

1. 快速检索 是 SinoMed 的默认检索方式, 其功能类似于搜索引擎, 用户无需选择检索入口, 系统默认在各数据库的全部可检字段中进行“智能检索”。如输入多个检索词, 中间用空格隔开, 系统默认对检索词进行“逻辑与”检索, 具体见图 4-2。

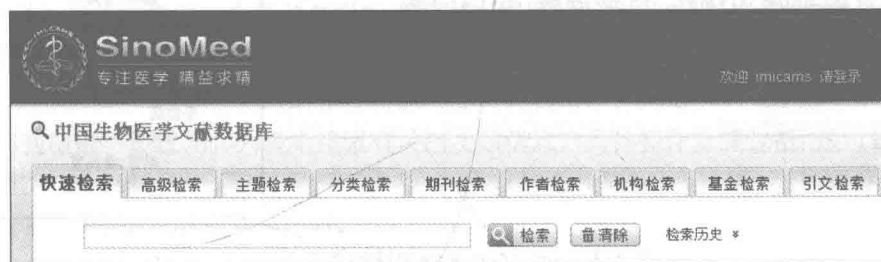


图 4-2 SinoMed 快速检索界面

2. 高级检索 采用自由词检索和字段检索结合的方式, 通过表达式构建器生成专业的检索表达式, 具体操作步骤如下:

(1) 根据需要选择检索入口: SinoMed 现提供常用字段、全部字段、中文标题、英文标题、摘要、关键词、主题词、作者、第一作者、作者单位、刊名、基金等 19 个检索入口, 其中“常用字段”是为方便同时在多个字段中进行检索而设置的, 是中文标题、摘要、关键词和主题词这 4 个常用检索项的组合。

(2) 在表达式构建输入框中输入检索词: 根据需要选择是否进行“智能检索”、“精确检索”, 并选择逻辑组配符, 点击“发送到检索框”按钮; 如需要构建多个检索词组配的表达式, 只需再次执行上述操作并选择合适的逻辑组配符即可; 对检索框中的检索表达式确认无误后, 点击“检索”按钮执行检索。

3. 智能检索 SinoMed 中, 智能检索指自动将用户输入的检索词转换成表达同一概念的一组词进行查询, “常用字段”“全部字段”“中文标题”“英文标题”“摘要”“关键词”均支持智能检索功能。如在“常用”字段输入“艾滋病”, 勾选“智能检索”执行检索后, 系统自动检出“中文标题”“摘要”“关键词”“主题词”中含“艾滋病”“AIDS”和“获得性免疫缺陷综合征”的所有文献。

4. 主题检索 指采取规范化的主题词基于主题概念进行检索。与“自由词检索”相比,“主题检索”能有效提高查全率和查准率。具体操作如下:

(1) 主题词定位: 在中 / 英文主题词检索入口键入完整检索词或片段, 查找浏览相关主题词注释信息和树形结构, 选择确定需要的恰当主题词, 如图 4-3 所示。

(2) 主题文献查找: 在选定主题词的注释信息显示界面, 选择是否加权、是否扩展, 添加相应副主题词后, 选择合适的逻辑组配符, 并点击“发送到检索框”按钮; 如需要进行多个主题词的组配检索, 只需要查找其他主题词并进行上述操作即可, 如图 4-4 所示。

(3) 扩展检索: 是对当前主题词及其下位主题词进行检索, 非扩展检索则仅限于对当前主题词进行检索, 系统默认为扩展检索。



图 4-3 SinoMed 主题词注释信息显示

(4) 加权检索: 表示仅对加星号(*)主题词(主要概念主题词)检索, 非加权检索表示对加星号和非加星号主题词(非主要概念主题词)均进行检索。系统默认为非加权检索。

(5) 副主题词组配检索: 指用副主题词来对主题词进行限定, 使检出的文献限于主题词概念的某一方面, 以提高检索的准确性。选择某一副主题词, 表示仅将当前主题词组配该副主题词的文献检出。如:“肝 / 药物作用”表明文章并非讨论肝脏的所有方面, 而是讨论药物对肝脏的影响。

(6) 副主题词扩展检索: 选择“扩展副主题词”, 指对当前副主题词及其下位副主题词进行检索; 非扩展检索则仅限于对当前副主题词进行检索。系统默认为副主题词扩展检索, 如图 4-4 所示。

5. 分类检索 从文献所属的学科角度进行检索, 具有族性检索的功能。检索入口包括分类导航、类号和类名, 可通过选择是否扩展、是否复分使检索结果更为贴切。具体操作如下:

(1) 在类名、类号检索入口输入学科类名或类号片段, 在“分类表”中查找浏览、选择合适的类名或类号; 也可通过分类导航逐级展开, 查找合适的类名。

(2) 在选定类名或类号的注释信息显示界面, 选择是否扩展, 添加相应的复分号后, 点击“分类检索”即可。“扩展检索”表示对该分类号及其全部下位类号进行检索, 不扩展则表示仅对该分类号进行检索。



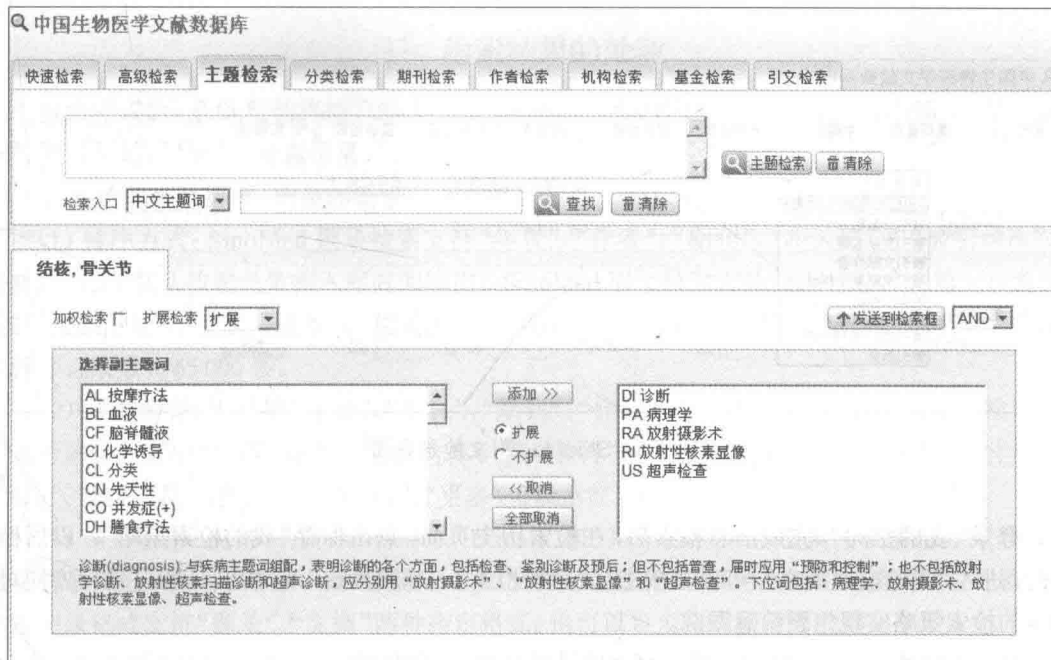


图 4-4 SinoMed 主题检索界面

(3) 复分组配检索：系统自动将能够与分类号组配的复分号列出，选择“全部复分”表示检索当前分类号与其中任何一个复分号组配及不组配任何复分号的文献；选择“无复分”表示检索当前分类号不组配任何复分号的文献；选择某一复分号表示仅检索当前分类号与该复分号组配的文献。

6. 期刊检索 通过期刊列表来检索有关的文献信息。具体操作步骤包括：通过检索入口选择刊名、出版地、出版单位、期刊主题词或者 ISSN 直接查找期刊；也可通过“期刊分类导航”或“首字母导航”逐级浏览查找期刊信息。

7. 作者检索 基于作者辅助表支持的“作者检索”，不仅可以实现“第一作者”检索，还可进行“作者机构限定”，提高作者检索的查准率与查全率。具体方法为进入“作者检索”界面，输入作者姓名，勾选“第一作者”后即指定为第一作者查找；根据实际需求对作者单位进行选择（可多选），点击“查找”按钮可得到选定机构中某作者发表的所有文献。

8. 机构检索 指通过文献的发文机构来检索相关的文献信息。SinoMed 数据库对发文机构进行了严格的形式规范，可以最大程度解决机构名称的“同构异名”问题，提高机构检索的查准率与查全率。具体方法为：通过“机构分类导航”或“首字母导航”逐级浏览查找机构，也可在检索框中输入机构名称直接查找某一机构。选择相应的机构名称，可点击“检索”按钮可得到选定机构发表的所有文献，点击“第一机构检索”或者“分析”按钮可做进一步限定或者多维度定量分析。

9. 基金检索 指通过文献的基金资助信息来检索相关的文献。SinoMed 数据库中的基金名称进行了形式规范，实现基金与项目的有效关联，提高基金检索的查准率与查全率。

10. 引文检索 引文检索是 SinoMed 新增的一项重要功能。支持从被引文献题名、主题、作者/第一作者、出处、机构/第一机构、资助基金等途径查找引文，帮助用户了解感兴趣文献在生物医学领域的引用情况（图 4-5）。此外，引文检索还提供了引文追踪和引文分析功能。引文追踪可以查看文献最近 5 年的被引次数，创建引文报告可提供总被引频次、篇均被引频次、h 指数等多维度的引文分析结果。

11. 定题检索 用于按照既定的检索策略定期跟踪某一课题的最新文献。首先需要免费注册

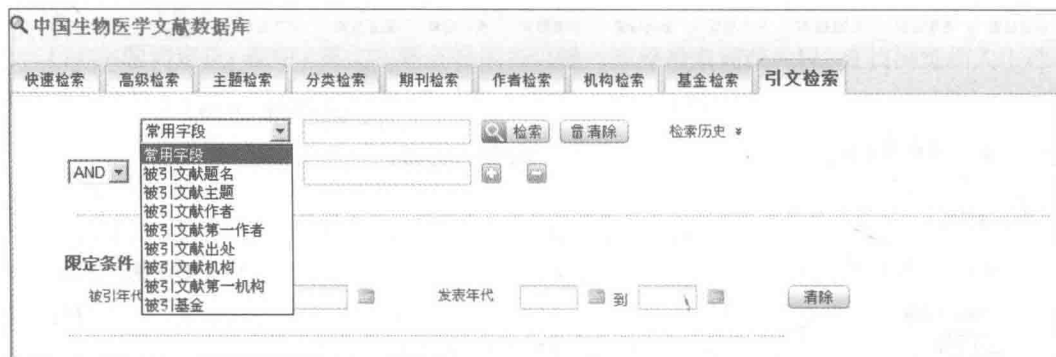


图 4-5 SinoMed 引文检索界面

册并登录“我的空间”，完成信息检索后，在检索历史页面，点击保存“我的检索策略”。以后检索时，进入“我的检索策略”，可以直接进行“重新检索”、“邮箱推送”或“RSS 订阅”，这样方便对相应的检索策略实现定题信息跟踪。

12. 链接检索 SinoMed 提供了强大的链接功能，主要包括：

作者链接：点击作者名，检索该作者发表的所有文献。

期刊链接：点击期刊名称，检索该期刊收录的所有文献；点击期刊卷期，检索该卷期收录的文献。

关键词链接：点击关键词，在缺省字段检索该词。

主题词、副主题词链接：点击主题词，检索该主题词标引的所有文献；点击副主题词，仅检索该主题词与副主题词组配标引的文献。

特征词链接：点击特征词，在特征词字段检索含该词的文献。

相关文献链接：点击“主题相关”，检出按内置算法判定的该文献的主题相关文献；点击“参考文献”，显示该篇文献的参考文献。

全文链接：点击全文链接图标，从维普科技期刊数据库中获取全文。

13. 检索方法的综合运用 面对一个检索任务，经常要综合运用多种检索方法，发挥不同检索方式的优势，检索结果才能更为全面、准确。实际上文献检索的过程也是科研过程的一部分，根据检索结果的情况调整检索策略和检索方式。比如一个检索任务“在 SinoMed 数据库中检索近 20 年来肝癌治疗方法的研究进展”，首先要分析此检索任务，可提取“肝癌”、“治疗”两个检索概念，为提高检出文献的相关性和全面性，首选“主题检索”途径，以“高级检索”中的标题自由词检索为补充，最后结合“限定检索”来进行数据时间范围与文献类型的限定。具体过程主要包括进入“主题检索”，在检索输入框中输入“肝癌”，点击“查找”按钮，系统显示主题词“肝肿瘤”为该词的同义词，且允许与副主题词“/ 治疗”组配。在此，选择“主题扩展”、“副主题扩展检索”，添加“/ 治疗”副主题词后点击“发送到检索框”按钮，最后点击“主题检索”按钮执行主题检索。然后进入“高级检索”，选择“中文标题”字段，在表达式构建器中分别输入“肝癌”、“治疗”，并选择合适的逻辑组配符，构造表达式“(((“肝癌”[中文标题：智能])OR “肝肿瘤”[中文标题：智能])OR “肝脏肿瘤”[中文标题：智能])AND “治疗”[中文标题：智能])”进行检索。最后进入“检索历史”，对前两步操作进行“逻辑或(OR)”运算，在检索结果显示页面进行限定操作，输入年代范围“1994—2014”，如果强调综述性文献，可勾选文献类型“综述”进行限定检索，也可根据需要进一步链接全文。

三、检索结果的处理

检索完成后,系统自动将检索结果分为“全部”、“核心期刊”、“中华医学会期刊”、“循证文献”,默认为显示“全部”检索结果。

1. 检索结果的排序、聚类 and 输出

(1) 排序方式: SinoMed 题录检索支持“年代”、“作者”、“期刊”、“相关度”和“被引频次”5种排序方式,默认按题录数据入库时间输出。SinoMed 引文检索支持“被引频次”、“第一作者”、“被引文献出处”、“文献发表年”、“相关度”5种排序方式,默认按被引频次输出。系统支持的最大排序记录数为 65 000 条。

(2) 结果聚类: 支持按“主题”、“学科”、“期刊”、“作者”、“时间”、“地区”6个维度对检索结果进行聚类。其中“主题”最多可展示到第 6 级;“学科”最多可展示到第 3 级;“期刊”、“作者”、“时间”、“地区”默认展示前 10 位,点击“更多”可展示前 50 位。

(3) 结果输出: 支持“打印”、“保存”、“E-mail”和“写作助手”4种检索结果输出方式,单次输出的最大记录数为 500 条。通过“标记记录”或“记录号”可选择感兴趣的某篇文献进行输出。输出记录格式支持“题录”、“文摘”两种指定格式,也可以自定义选择需要的输出字段。

2. 查看相关文献 SinoMed 中相关文献的类型包括“主题相关”、“共引相关”、“作者相关”3种类型。在检索结果页面可以直接点击相应按钮进行链接检索;也可以通过点击相应文章的标题,进入文章详细页面后,直接浏览该文献的相关文献。

3. 检索历史的查看、编辑 系统将每次的检索步骤记录在“检索历史”中,包括序号、检索表达式、命中文献数、检索时间和推送方式。推送方式包括邮箱推送和 RSS 订阅两种方式;检索历史中最多保存最近检索的 200 条检索表达式,按照时间顺序从上往下依次显示,最新的检索式总在最上方。

在“检索历史”中可对检索记录进行逻辑组配操作。可从中选择一个或多个检索表达式并用逻辑运算符 AND、OR、NOT 组成更恰当的检索策略;检索策略可以永久保存到系统提供的“我的空间”。

四、个性化服务

通过 SinoMed 的“我的数据库”,可对感兴趣的文献进行在线保存、管理与再利用。注册并登录“我的空间”后,在检索结果界面点击“我的数据库”便可将当前文献题录添加到在线数据库中;通过标签和备注,可以从课题、项目、学科、主题等角度对收藏的文献进行分类组织和标注;通过设置的“标题”、“作者”和“标签”检索入口,可方便对历史存储文献进行检索和再利用。

第二节 中国医院知识总库(CKHD)

一、概述

中国医院知识总库(China Hospital Knowledge Database, CHKD)是清华同方向医药领域信息需求的数据库产品。CHKD 总库包含有《CHKD 期刊全文数据库》《CHKD 博硕士学位论文全文数据库》《CHKD 会议论文全文数据库》《CHKD 报纸全文数据库》《CHKD 国家科技成果数据库》《CHKD 年鉴网络出版总库》《CHKD 专利全文数据库》《CHKD 标准数据库》《人民军医图书数据库》等十一类资源。其中《CHKD 期刊全文数据库》收录自 1915 年至今我国公开出版发行的医学类期刊 1534 种,医学相关期刊 5600 多种(收录部分期刊内容),独家授权期刊共 415 种,累计收录文献量超过 1000 万篇。



作为中国知识基础设施工程(China National Knowledge Infrastructure, CNKI)旗下的子数据库,CHKD 收录文献以医学为主,涉及基础医学、临床医学、预防医学、中医学、中药学、信息管理、卫生政策研究等各领域,CHKD 提供 WEB 版(网上包库)、镜像站版、光盘版、流量计费等多种服务方式,网络版内容每日更新。网络版的访问网址是: <http://www.chkdcnki.net/>。

CHKD 收录期刊、学位论文、会议论文、报纸的特点及范围见表 4-1。

表 4-1 CHKD 收录各类文献的特点及范围

类型	电子期刊	学位论文	会议论文	报纸
收录年限	1915 年至今	1984 年至今	1999 年至今	2000 年至今
收录内容	生物医学类期刊和相关专业期刊	全国博士培养单位的优秀博、硕士学位论文。独家或唯一博、硕士论文供稿单位 238 家	中华医学会等权威学会在国内会议上发表的医药卫生类会议论文	重要报纸刊载的学术性、资料性文献
数量	近 1000 万篇	近 37 万篇	6100 种会议论文集,年更新约 9 万篇文章	医药卫生类报纸 15 种,相关报纸 1000 多种,年更新 6 万篇

二、检索方法和技巧

通过统一的检索界面,可对期刊、论文、会议、报纸多个数据库进行跨库检索及导航浏览检索。下面以 CHKD 全文期刊为例介绍 CHKD 主要的导航浏览和检索方法。

(一) 资源导航

CHKD 检索首页上方显示“资源导航”栏,包括医学学科馆、中华预防医学会期刊、中国药学会期刊、人民军医图书等栏目。点击“资源导航”标签,可展开“期刊导航”“博硕导航”“年鉴导航”。

1. 分类导航和知识导航 在 CHKD 单库检索的页面,系统提供分类导航和知识导航功能(图 4-6)。分类导航是中国图书分类法的导航,点击分类名称可了解某一个学科类目下文献的相关情况。知识导航即主题词导航,可点击知识导航相关主题,逐级深入即可检索和浏览。分类导航和知识导航可以同时进行选择和浏览,分类和知识导航间组配的逻辑关系是“与”,导航内组配的逻辑关系是“或”。

2. 期刊导航 在 CHKD 全文期刊数据库页面点击“期刊导航”按钮,可从期刊刊名首字母、期刊类别对所有收录期刊进行导航浏览。如点击“核心期刊导航”“独家授权期刊导航”等即显示相应各类的期刊,并进一步显示该刊某一卷期的文献内容。

(二) 检索

1. 快速检索 在 CHKD 首页提供快速检索的检索方式。快速检索类似搜索引擎,只需在检索框中输入所要找的关键词,点击“检索”就可以查到相关的文献(图 4-7)。

2. 标准检索 点击 CHKD 首页的检索输入框右上方的“高级检索”,可进入到提供多种检索方式(包括标准检索、主题词检索、检索式检索、科研基金检索等)的页面。在标准检索中,可先输入检索相关的常用限定条件如发表时间、文献来源(期刊、报纸、年鉴或博硕士学位授予机构)、作者、作者机构等,再选择检索项(检索字段)输入检索词。检索词可限定在“题名 & 关键词 & 摘要”、“题名”、“关键词”、“摘要”、“全文”、“参考文献”6 个检索字段。检索词间可使用“并且”、“或者”、“不包含”三种方式进行逻辑关系的组合检索。点击检索项前的  增加逻辑检索行,添加另一个文献内容特征检索项;点击  减少逻辑检索行。CHKD 提供了灵活多样的检索限定,包括检索用词出现的次数(频率)、检索年代范围、期刊范围、是否精确匹配等(图 4-8)。





图 4-6 CHKD 分类导航和知识导航界面

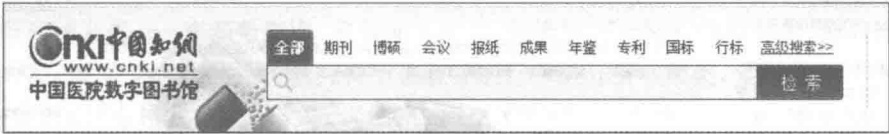


图 4-7 CHKD 快速检索界面

“词频”指输入的检索词在所选择的检索项(即检索字段)中出现的频次。词频为空,表示在选择的检索项中至少出现 1 次检索词;如果为词频数字“3”,则表示在检索项中该检索词至少出现 3 次。

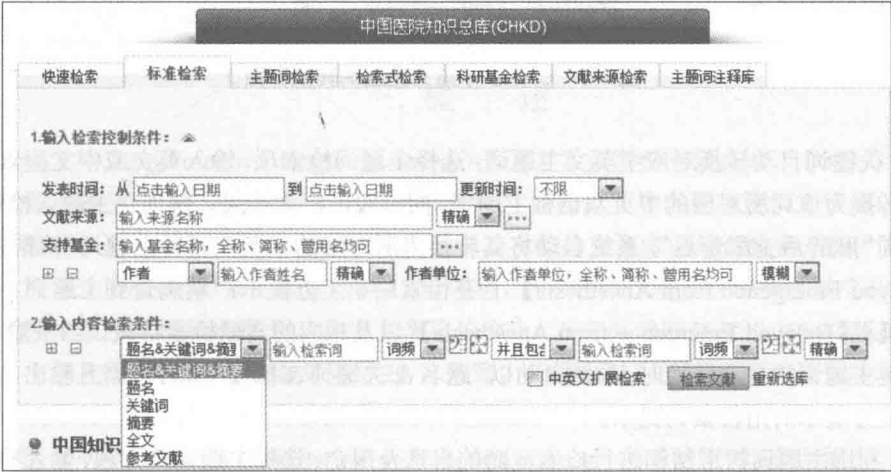


图 4-8 CHKD 标准检索界面

“最近词”点击图标,弹出的窗口显示本次登录最近输入的10个检索词。点击其中某一检索词,则该检索词自动进入检索框中。

“词扩展”点击图标,弹出的窗口显示以输入词为中心的相关词。相关词可以单词自动增加、多词自动增加、相关词取代原输入词等方式添加到检索框中。

精确匹配是指检索结果完全等同或包含与检索字/词完全相同的词语;模糊匹配是指检索结果包含检索字/词或检索词的片段。

3. 主题词检索 CHKD 数据库提供强大的主题词检索功能,可利用主题词树形结构进行检索范畴的自选及限定,并实现关键词自动转化为主题词功能。主要包括:

(1) 中英文主题词相互自动转换:输入英文或中文主题词,系统便自动转换为检索项设定的主题词,同时输出检索结果。例如在主题词检索项中输入英文主题词“Uterine Artery Embolization”,系统自动将其转换为中英双语种主题词“子宫动脉栓塞术【Uterine Artery Embolization】”,同时输出检索结果。点击“导航定位”左侧显示主题词“子宫动脉栓塞术”树形结构,表明“子宫动脉栓塞术”归属“栓塞,治疗性”和“妇科外科手术”概念范畴(图4-9)。



CHKD 期刊全文数据库 文献检索 期刊导航

分类导航: 全选 清除 总目录

知识导航: 全选 清除 总目录

医学、卫生

西医学、西药学

分析、诊断和治疗技术及设备

外科手术 → 本位词

泌尿生殖外科手术 → 本位词

妇科外科手术 → 本位词

子宫动脉栓塞术

治疗学 → 本位词

治疗性闭塞 → 本位词

栓塞,治疗性 → 本位词

子宫动脉栓塞术

止血技术 → 本位词

栓塞,治疗性 → 本位词

子宫动脉栓塞术

检索历史: 词表在工具书中的解释如下: 当前检索词的相似词: 当前检索词的相关词: 当前主题词在外文库中检索:

PubMed

快速检索 标准检索 主题词检索 检索式检索 科研基金检索 来源期刊检索 主题词注释库

输入中文或英文检索词: 主题词 Uterine Artery Embolization 精确

并且 题名&关键词&摘要 输入检索词 并且包含 输入检索词 精确

发表时间: 从 点击输入日期 到 点击输入日期

中英文扩展检索 在结果中检索 检索文献

精确查到主题词1个: 导航定位

子宫动脉栓塞术【Uterine Artery Embolization】

定制本次检索式

您可以按如下文献分组排序方式选择文献: (分组只对前4万条记录分组,排序只在800万条记录以内有效)

文献分组浏览: 主题词类别 学科类别 期刊名称 研究资助基金 研究层次 文献作者 作者单位 中文关键词 发表年度 不分组

文献排序浏览: 发表时间 相关度 被引频次 下载次数

每页记录数: 10 20 50

摘要显示 列表显示

共有记录2007条 首页 上页 下页 末页 1 / 141 翻页

序号	题名	作者	刊名	年期	被引频次	下载次数
1	剖宫产瘢痕妊娠47例临床诊疗分析	廖慧芬	内蒙古医学杂志	2014/04		
2	子宫动脉栓塞术治疗早期宫外孕的临床疗效观察	刘建梅,朱静,程湘,杨丽	局解手术学杂志	2014/02		
3	介入治疗在中孕妊娠前置状态的应用	屈萍,李维珍	吉林医学	2014/11		
4	子宫动脉栓塞术治疗瘢痕子宫合并前置胎盘疗效观察	樊阳阳,折开雄	现代中西医结合杂志	2014/11		
5	子宫动脉栓塞术与子宫切除术治疗产后出血的临床疗效对比分析	杨丽	中国卫生产业	2014/10		

图4-9 CHKD 中英文主题词转换界面

(2) 关键词自动转换对应中英文主题词:选择主题词检索项,输入英文或中文检索词,系统可自动转换为该词所对应的中英双语种主题词,同时输出检索结果。例如在主题词检索项中输入检索词“麻醉后觉醒延迟”,系统自动将其转换为所对应的中英双语种主题词“麻醉后苏醒延迟【Delayed Emergence from Anesthesia】”且在检索结果上方提示:“精确查到主题词1个:麻醉后苏醒延迟【Delayed Emergence from Anesthesia】”以及相应的文献检索结果。当所输入检索词无法实现主题词的自动转换时,系统自动以“题名&关键词&摘要”进行检索且输出结果,同时在结果上方显示所用检索词提示。

(3) 利用主题词树形结构进行检索范畴的自选及限定:选择主题词检索项,输入检索词,输出检索结果,点击导航定位,左侧知识导航显示该主题词的树形结构,包括:主题词、本位词、上



位词、同位词和下位词。利用主题词树形结构可进行主题词检索、本位词检索、上位词检索、同位词检索、下位词检索,来自选及限定检索范畴。

(4) 主题词模糊查询:选择主题词检索项,输入中文检索词,匹配中选择“模糊”,系统可自动对应出全部含有所输入检索词词素的中英双语种主题词,同时输出首词检索结果。例如:在主题词检索项中输入中文检索词“巨核细胞”,模糊查询结果为含有“巨核细胞”词素的“巨核细胞-红系祖细胞【Megakaryocyte Progenitor Cells】”等三个中英双语种主题词,且以词首字汉语拼音顺序横向排序。页面显示的是首词(即排序为首位中英双语种主题词“巨核细胞【Megakaryocytes】”)的检索结果。

三、检索结果的处理

1. 知网节扩展检索 在检索结果页面上点击每一文献题名,进入当前篇名文献细览区,即知网节,提供单篇文献的详细信息和扩展信息浏览。知网节包含了单篇文献的详细信息即:题名、作者、作者单位、文献出处、中英文关键词、摘要、中英文主题词等细览、全文快照搜索。

(1) 知网节提供中/英文关键词、中/英文主题词链接功能:将鼠标放置在文献细览界面的任一中/英文关键词或中/英文主题词时,便可获取该词汇在“知识元”、“工具书”、“文献检索”中的相关信息,并可查看该关键词或主题词的全部文献且以所限定年限倒序排序输出。

(2) 知网节中还提供以节点文献为中心的知识网络:可以看到所引用参考文献的记录、被引用情况及相关文献的记录,具体包括引文网络、参考文献、共引文献、引证文献、相似文献、同行关注文献、相关作者文献、相关机构文献、文献分类导航、文献分别来自于哪个数据库等各种扩展信息。

2. 检索结果的分组、排序和导出 CHKD 中的检索结果可按照相关度、发表时间、被引频次和下载频次进行排序,可快速筛选出最相关、最新、高质量及高关注度文献。此外检索结果还可按照多种方式——来源数据库、主题词类别、学科类别、关键词、文献作者、作者单位、文献出版来源、发表年度、研究层次、研究资助基金等进行分组分析与聚类,实现知识挖掘。CHKD 提供九种题录保存格式:简单格式、详细格式、引文格式、自定义格式、Refworks、EndNote、NoteExpress、查新格式、同方知网 pdl。

3. 全文下载 CHKD 提供 CAJ 和 PDF 两种全文格式。点击“CAJ 下载”或“PDF 下载”,根据系统提示选择打开或保存(注:CAJ 格式文件需用同方独有全文浏览器 CAJ Viewer 进行阅读)。

第三节 万方医学网

一、概述

万方医学网(<http://med.wanfangdata.com.cn/>)是万方数据股份有限公司于 2009 年推出的医药信息专业服务平台,收录生物医学期刊、学位论文、会议论文等多种类型中文资源的全文,同时也提供来自 PubMed、OA 期刊、NSTL 生物医学论文的外文资源的整合检索。在中文期刊上,万方医学网收录 1998 年以来的医学期刊 1100 多种,其中包括 200 多种中华医学会和中华医师协会独家授权数字化出版的期刊。在中文学位论文上,万方医学网包含了 1999 年以来博硕士学位论文文摘 35 余万条,学位论文全文 22 万余篇。此外通过万方医学网还提供医学视频、临床诊疗知识库、医学图书、中医药知识系统等特色资源。

万方医学网是万方数据(<http://www.wanfangdata.com.cn/>)的子数据库,其中的期刊、学位论文和会议论文同时也可以通过万方数据平台获取全文。

二、检索方法和技巧

(一) 资源导航

主页上提供了期刊导航、学位导航、会议导航、中华医学会期刊导航等导航方式。期刊导航可以按学科分类、国外数据库收录、国内核心期刊收录、期刊主管单位、刊名首字母等方式浏览,如图4-10所示。



图 4-10 万方医学网期刊导航界面

(二) 检索

1. 快速检索 主页的上方可直接进行快速检索。快速检索的默认状态是 unlimited 检索字段同时检索期刊、学位论文和会议论文。点击快速检索框上不同的限定标签,即可限定在不同的字段和范围内进行检索。例如点击“期刊”标签,在检索框输入“消化”,即在期刊刊名字段进行检索,检索结果是刊名中包括了“消化”的期刊;点击“作者”标签,在检索框输入“李芳”,即是在作者字段进行检索,检索结果可以浏览包含了“李芳”的不同作者名称及其对应的机构、发表论文数量。

2. 高级检索 高级检索帮助进行更精确的检索,可输入多个检索词,并将检索词限定在不同的检索字段(检索项),并进行检索词间的“与”、“或”、“非”的逻辑组配。在主页上方快捷搜索的右侧,点击高级搜索,即可进行高级检索。高级检索分为跨库检索、期刊检索、外文检索、学位检索、会议检索、Mesh 主题检索 6 个检索模块。6 个模块的检索是相同的,这里以跨库检索为例进行说明。

跨库检索中输入的检索词可限定在“标题”、“关键词”、“作者”、“摘要”、“药品”、“疾病”等 16 个检索字段。例如查找 2013 年以来发表的如何预防羊水栓塞的相关文献,可将检索字段限定为摘要,检索词的关系为“与”(图 4-11)。点击“”,可增加检索项。也可在检索栏的下方进行限定,限定在期刊论文或者学位论文等,进一步限定特定类别的文献信息。



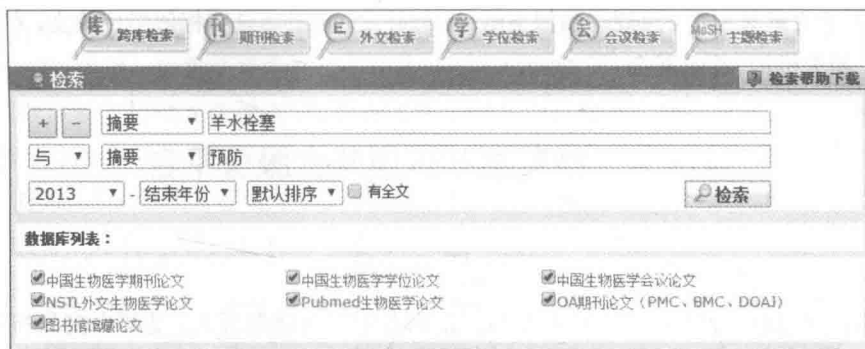


图 4-11 万方医学网跨库检索界面

三、检索结果的处理

1. 检索结果的排序、聚类 and 导出 检索结果提供按时间、相关度和被引次数 3 种排序方式。同时在检索结果页面的左侧按照论文类型、分类、刊名、时间进行了聚类统计,可直观的浏览检索结果的时间分布、期刊分布(图 4-12)。



图 4-12 万方医学网检索结果聚类界面

2. 知识点链接 检索到的每篇文献不仅提供了文献作者、关键词、期刊名、卷期、摘要等基本信息,还提供了是否核心期刊、被引次数等。标识 ISTIC 表示期刊被《中国科技论文统计源期刊》收录,标识 PKU 表示期刊被《中文核心期刊要目总览》(北大核心期刊)收录。通过点击文献的作者、关键词、期刊名、卷期,可直接检索和跳转到相应的检索结果页面。

点击文献的篇名,进入文献的详细信息页面。详细信息页面提供了更多的知识点链接,页

面下方提供了文献被引用的列表、相似文献、相似外文、相似会议和相似学位论文,右侧可浏览相关作者、相关机构、相关基金。

第四节 维普期刊资源整合服务平台

一、概述

维普期刊资源整合服务平台由重庆维普资讯有限公司研制开发,是其中文科技期刊数据库的升级版。整个平台收录了1989年以来国内公开出版的12000余种期刊,含中文核心期刊1980余种、近4000万篇科技文献,年增文献约300万篇。学科范围涵盖社会科学、自然科学各个领域。访问网址为<http://lib.cqvip.com/>,网络版数据每日更新。各个机构根据订购学科包的不同,包含的期刊、科技文献的数量可能不同,具体可以向所在机构的图书馆员咨询。2006年开始为普中文科技期刊全文数据库与Google Scholar学术搜索合作服务,通过Google Scholar可直接检索中文科技期刊全文数据库的题录信息。

维普期刊资源整合服务平台不仅提供原始文献信息服务,而且更多地提供深层次知识服务。平台分为期刊文献检索、文献引证追踪、科学指标分析、高被引析出文献、搜索引擎服务5个功能版块。“期刊文献检索”可检索到文献信息,提供在线阅读、下载全文;“文献引证追踪”的功能是追踪和揭示文献相互引用关系;“科学指标分析”辅助分析学科热点和研究绩效;“高被引析出文献”可基于期刊参考文献析出其他类型高被引文献资源;“搜索引擎服务”可通过谷歌、百度等搜索引擎扩展检索(图4-13)。



图4-13 维普平台期刊文献检索界面

二、检索方法和技巧

这里以“期刊文献检索”模块为例,介绍检索方法和技巧。“期刊文献检索”提供基本检索、传统检索、高级检索3种检索方式及期刊导航、检索历史功能。

1. 基本检索 在基本检索中,可先选择检索时间、期刊范围(全部期刊、核心期刊、SCI来源期刊等)、学科,然后选择检索字段项,输入检索词,选择检索关系(与、或、非),进行检索。

2. 传统检索

(1) 限定范围:系统提供学科类别、期刊类别、数据年限和期刊范围的限制。可根据《中国图书馆分类法》(第四版)的分类体系对检索范围进行学科类别限制,或者限定某一类期刊,如核心期刊、重要期刊、SCI期刊、CA期刊等。

(2) 辅助检索功能

① 同义词:如选择“同义词”功能,系统自动扩展检索所有同义词,以提高检全率。该功能适用于关键词、题名或关键词、题名三个检索字段。



②同名作者: 如选择“同名作者”功能, 检索某一作者发表文献时, 系统首先显示该姓名作者的单位列表, 供进一步确认后选择。该功能可用于“作者”、“第一作者”字段。

3. 高级检索 提供 5 个检索项, 可输入检索词并进行检索字段的限定。与基本检索不同的, 高级检索中提供了查看同义词、同名/合著作者、查看相关机构、查看分类表、期刊导航等辅助功能。例如在“题名或关键词”字段后的检索框, 输入“子宫肌瘤”, 点击“查看同义词”, 可浏览与子宫肌瘤相关或同义词, 同时可勾选需要的词语, 加入到检索栏内(图 4-14)。

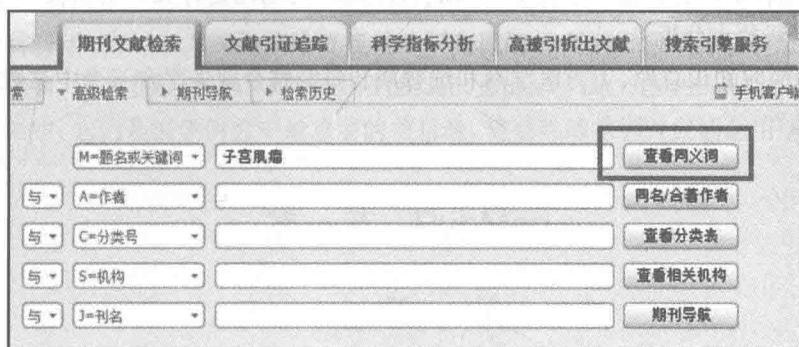


图 4-14 展示了维普平台期刊文献检索模块的高级检索界面。界面顶部有五个主要功能标签：期刊文献检索、文献引证追踪、科学指标分析、高被引析出文献、搜索引擎服务。下方是高级检索的入口，包含“高级检索”、“期刊导航”和“检索历史”三个子标签。在“高级检索”子标签下，有一个“M=题名或关键词”的检索框，其中输入了“子宫肌瘤”。该检索框右侧有一个“查看同义词”按钮。下方还有四个其他检索项，每个都包含一个“与”字样的下拉菜单、一个检索框和一个辅助功能按钮：A=作者（同名/合著作者）、C=分类号（查看分类表）、S=机构（查看相关机构）、J=刊名（期刊导航）。

图 4-14 维普平台期刊文献检索模块高级检索界面

4. 期刊导航 直接输入刊名、按照刊名第一个字首字母、学科分类、核心期刊、国外数据库收录、期刊地区分布六种期刊导航方式。

5. 检索历史 可浏览检索历史。通过“检索历史”, 可进行检索式的逻辑“与”、“或”、“非”的组配。

三、检索结果的处理

“期刊文献检索”模块中的检索结果按相关性排序。如有全文的文献会标识“在线试读”和“下载全文”, 仅有文摘的文献会标识“文献传递”。“文献引证追踪”模块中检索结果按被引次数排序。

本章小结

本章选取了查找中文信息资源常用的四个数据库, 既包括文摘数据库中国生物医学文献服务系统 SinoMed, 也包括三个中文全文数据库中国医院知识总库、万方医学网和维普期刊资源整合服务平台。严格来讲 SinoMed 已经不单纯是一个中文期刊文摘数据库, 它还包括外文(英、日、俄)医学期刊论文以及开放获取全文、协和医大学位论文等, 但是基于主题词标引的中文期刊文摘检索依然是该数据库的特色。本章中的三个中文期刊全文数据库收录内容和检索功能各有特色, 相互交叉, 但都不能全部涵盖国内医学期刊内容, 提醒大家检索时应该检索多个数据库, 最后用文献管理软件或者其他工具来去重, 才能保证中文文献检索结果尽量全面。

(任慧玲)

思考题

1. 中国生物医学文献服务系统 SinoMed 包含哪些资源内容?
2. SinoMed 提高查全率的方法有哪些? 提高查准率的方法有哪些?

3. SinoMed 检索结果页面提供了哪些链接信息?
4. CHKD 的跨库检索都包含哪些类型的文献? 哪些资源是 CHKD 独家收录的?
5. CHKD 知网节可获取哪些扩展的信息?
6. 万方医学网包含哪些特色信息内容?
7. 万方医学网知识点链接可以直接检索什么信息?
8. 维普期刊资源整合服务平台主要包含哪五个功能模块?
9. 维普期刊资源整合服务平台中“文献引证追踪”的功能是什么? 请以某一个课题为例, 练习引文和学科热点分析。
10. 中国医院知识总库、万方医学网和维普期刊资源整合服务平台三个中文数据库的特色和区别是什么?



第五章 外文文摘数据库检索

在医学科研活动中,要系统全面地查找某一研究课题的文献信息,首先需要从文摘数据库中获取文献线索,进而获取全文。本章介绍 PubMed、Embase、BIOSIS Previews、SciFinder 四种外文文摘数据库,它们是医学研究中最重要信息源,需要熟练掌握并灵活运用这些数据库。

第一节 PubMed

一、概述

PubMed 是生物医学领域最重要也最权威的数据库之一,由美国国立医学图书馆(NLM)下属的国家生物技术信息中心(NCBI)研制开发,可通过互联网免费访问。

PubMed 最早源于 1879 年创刊的医学检索工具 Index Medicus(IM)。20 世纪 60 年代,为实现 IM 的自动化编辑,NLM 创建了 MEDLARS 系统(Medical Literature Analysis and Retrieval System),其后又进一步实现了联机检索,称为 MEDLINE(MEDLARS Online)。1983 年,MEDLINE 发行光盘版并得到广泛应用,成为医学文献检索的最常用数据库。1996 年,NCBI 推出了基于 Web、以 MEDLINE 数据库为核心内容的 PubMed 检索系统,并免费向全球开放。PubMed 文献报道速度快,数据每天更新,检索功能强大,使用方便快捷,还提供丰富的外部链接及多种个性化服务功能,是当今移动互联网环境下全世界生物医学研究中不可或缺的文献信息资源。

1. PubMed 的收录范围及来源 PubMed 收录了来自全世界 80 多个国家和地区的 5600 多种生物医学期刊及部分在线图书的摘要信息,截至 2014 年 9 月,累计文献记录超过 2400 万条,绝大部分期刊文献可回溯至 1946 年,最早可回溯至 19 世纪初。PubMed 的文献记录主要有以下几类:

(1) MEDLINE: 为 PubMed 的主体内容,占总文献量的 90%,现有近 2200 万条文献记录。MEDLINE 的每条文献记录都进行了深加工,按照医学主题词表(MeSH)标引了 MeSH 主题词(MeSH Terms),还提供文献类型(Publication Types)、基因库索取号(GenBank Accession Numbers)等信息。这些记录标注为 [PubMed-indexed for MEDLINE]。

MEDLINE 数据库也整合在 Ovid、Web of Science、EBSCO、SciFinder、OCLC FirstSearch 等多个不同类型的数据库平台上,并与这些平台上的其他数据库实现跨库检索。

(2) in-process citations: PubMed 将尚未完成主题词标引等加工处理的最新文献存入临时库中,并标注为 [PubMed-in process]。

(3) publisher-supplied citations: 为加快报道速度,PubMed 同步发布出版商提供的电子文献信息,并标记为 [PubMed-as supplied by publisher]。大部分属于 MEDLINE 收录范围的记录随后会转为 [PubMed-in process] 状态,少部分不属于 MEDLINE 收录范围的记录(例如综合性期刊中的非生物医学文献),PubMed 会继续保留,其记录状态不变或仅标注为 [PubMed]。

(4) OLDMEDLINE: PubMed 包含了部分 1966 年以前出版且尚未被 MEDLINE 收录的文献记录,并标注为 [PubMed-OLDMEDLINE]。

(5) 其他：包含部分由美国国立卫生研究院(NIH)资助的研究人员提交的作者手稿(author manuscript)、NCBI 在线图书检索系统 Bookshelf 的电子书及研究报告，以及 NCBI 的生物医学期刊开放存取仓储平台——PubMed Central(PMC)收录的期刊文献等信息。PMC 目前收录了来自 2200 多种期刊的 300 多万篇全文记录，PubMed 中这些可以链接到 PMC 并免费获取全文的记录标注为“Free PMC Article”。此外，部分由出版商免费提供全文的记录，PubMed 则直接标注为“Free Article”。

PubMed 中每条记录都有唯一的识别号 PMID(PubMed Unique Identifier)。对出版商通过在线平台优先发布的电子期刊文献，PubMed 标注为 [Epub ahead of print]，并提供每篇文献的 DOI(数字对象唯一标识符)及全文链接地址。

PubMed 整合在 NCBI 的统一检索平台上，与该平台上的其他 40 多个数据库建立了无缝链接，全面提供了生物医学研究必需的文献信息与分子生物信息，包括文献库(Literature)、健康库(Health)、基因库(Genes)、基因组库(Genomes)、蛋白质库(Proteins)以及化学物质库(Chemicals)等六大类。这些数据库均可免费访问，并实现跨库检索。

2. PubMed 的常用检索字段 PubMed 的记录字段有 60 多个，可检索的字段有 43 个，表 5-1 列出了常用检索字段的标识(Tags)、字段名称及字段含义。

表 5-1 PubMed 的常用检索字段

字段名称	字段标识	字段含义
Affiliation	[AD]	第一著者的单位、地址(包括 Email 地址)
Article Identifier	[AID]	文献识别码，例如 DOI
All Fields	[ALL]	所有字段
Author	[AU]	著者
Corporate Author	[CN]	合作者或团体著者
EC/RN Number	[RN]	美国 FDA 物质登记系统唯一识别码，或国际酶学委员会给特定酶的编号，或化学物质的 CAS 登记号
Entrez Date	[EDAT]	文献被 PubMed 收录的日期
First Author Name	[1AU]	第一著者
Full Author Name	[FAU]	著者全名
Grant Number	[GN]	获资助项目的编号或合同号
Issue	[IP]	期刊的期号
Investigator	[IR]	对研究项目有贡献的主要研究者(PI)或合作者
Journal Title	[TA]	期刊全称、缩写或 ISSN 号
Language	[LA]	语种
Last Author	[LASTAU]	排名最后的作者
MeSH Major Topic	[MAJR]	主要 MeSH 主题词，主题词后加“*”作为标记
MeSH Subheadings	[SH]	MeSH 副主题词
MeSH Terms	[MH]	MeSH 主题词
Pagination	[PG]	文献在期刊中的起始页码
Place of Publications	[PL]	期刊的出版国别
PMID	[PMID]	PubMed 中文献的唯一识别码(Unique Identifier)
Publications Date	[DP]	文献的出版日期
Publication Type	[PT]	文献类型
Secondary Source ID	[SI]	第二来源库 ID(在 NCBI 其他数据库的识别码)

续表

字段名称	字段标识	字段含义
Supplementary Concept	[NM]	补充概念, 包括化学物质、方案、疾病术语等
Subset	[SB]	PubMed 数据库子集
Text Words	[TW]	文本词, 来自 TI、AB、MH、SH、PT、NM 等字段
Title	[TI]	文献的题名
Title/Abstract	[TIAB]	文献的题名 / 摘要
Volume	[VI]	期刊的卷号

二、检索方法与技巧

(一) 检索技术

1. 自动词语匹配检索 自动词语匹配 (automatic term mapping) 是 PubMed 的特色检索技术, 其基本原理是: 系统自动对输入的检索词进行概念分析, 在多个索引词表 (包括 MeSH 转换表、刊名转换表、著者索引及转换表、研究者索引及转换表) 中搜索、比对、匹配, 并转换为相应的 MeSH 主题词、刊名、著者或研究者, 再将检索词在所有字段 [All Fields] 中检索, 最后执行“OR”布尔逻辑运算。如果检索词是短语词组, 系统会将其拆分为单词后在所有字段检索, 单词之间的布尔逻辑关系为“AND”。检索结果显示页面右下方的 Search Details 栏目会显示 PubMed 实际执行的检索式。

例如, 输入 aspirin heart attack, Search Details 显示的检索式为: ("aspirin" [MeSH Terms] OR "aspirin" [All Fields]) AND ("myocardial infarction" [MeSH Terms] OR ("myocardial" [All Fields] AND "infarction" [All Fields]) OR "myocardial infarction" [All Fields] OR ("heart" [All Fields] AND "attack" [All Fields]) OR "heart attack" [All Fields])。

可以看出, PubMed 自动将 heart、attack 及 aspirin 这三个词确定为 heart attack (心脏病发作) 与 aspirin (阿司匹林) 两个概念, 并转换为相应的 MeSH 主题词 myocardial infarction 与 aspirin, 同时将检索词拆分后在所有字段检索, 再进行逻辑组配。因此, 自动词语匹配是一种智能化的检索过程, 能保证较好的查全率。

2. 布尔逻辑检索 PubMed 支持布尔逻辑检索, 运算符 AND、OR、NOT 分别表示逻辑“与”、逻辑“或”、逻辑“非”, 运算规则是按照检索式自左向右顺序运算, 圆括号为优先运算符, 可改变运算顺序。

3. 截词检索 PubMed 支持右截词检索, 以提高查全率, 截词符为星号“*”。截词检索时, PubMed 关闭自动词语匹配功能。

4. 短语检索 PubMed 的短语检索是将短语加上双引号进行精确检索, 含有连词符的短语, 或者将短语限定在特定字段如 [tw] (文本字段) 也可以进行精确检索。精确检索时, PubMed 关闭自动词语匹配功能, 不进行短语拆分。

5. 限定字段检索 PubMed 支持限定字段检索, 格式为: 检索词 [字段标识], 常用可检索字段及标识参见表 5-1。例如, hypertension [ti]、clinical trial [pt]、Lewis B [au] 分别表示检索篇名中含有 hypertension 文献、文献类型为 clinical trial 的文献以及作者为 Lewis B 的文献。

(二) 检索方法

PubMed 的检索方法包括基本检索、高级检索、主题词检索及专项检索。

PubMed 主页面 (图 5-1) 上方为检索区, 页面中部为 PubMed 的 3 个专栏, 分别是 Using PubMed、PubMed Tools、More Resources。页面底部是 NCBI 资源总览及帮助系统汇总。

1. 基本检索 进入 PubMed 主页, 默认为基本检索, 可在检索框中直接输入有实际意义的检索词, 如关键词、著者、刊名等, 点击“Search”, 系统执行自动词语匹配检索, 返回检索结果。



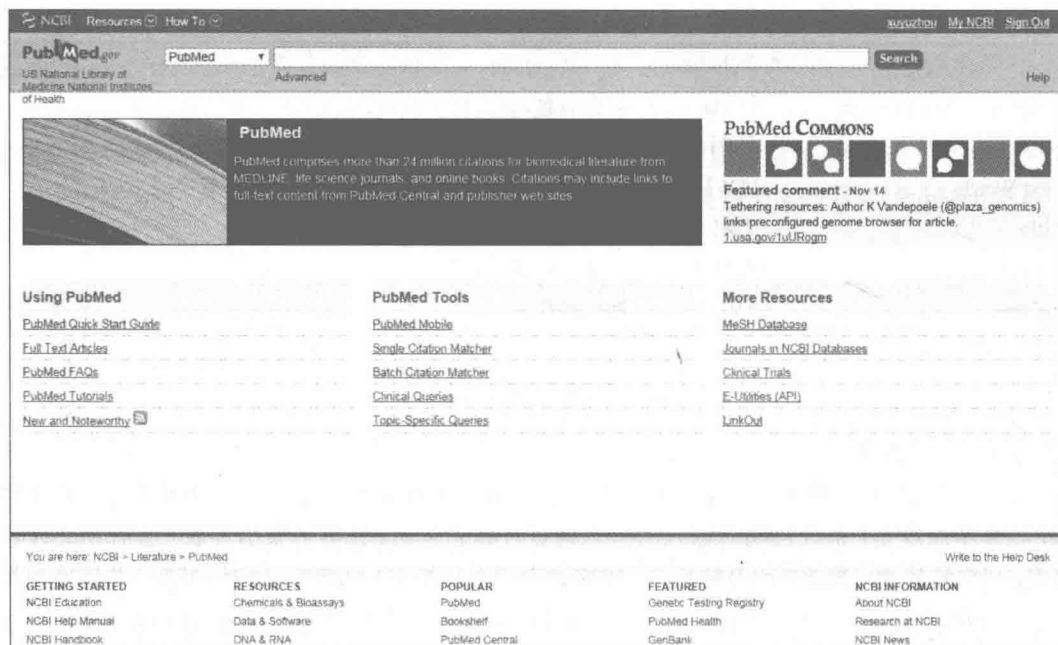


图 5-1 PubMed 主页面

输入检索词时, PubMed 有智能拼写检查及词语自动提示功能, 帮助用户正确选词。点击检索框左侧的 PubMed 下拉菜单, 可以选择 All Databases 或 NCBI 的其他数据库进行检索。

PubMed 的自动词语匹配检索功能强大, 简单好用, 能满足一般的查询需求。基本检索还可以通过著者检索、期刊检索、短语检索、截词检索、限定字段检索、布尔逻辑检索等途径, 实现复杂课题的检索。

(1) 著者检索: PubMed 自动执行前方一致的截词检索。例如, 输入 Smith M, 系统自动检索出 Smith MB, Smith MR 等所有姓为 Smith, 名字首字母为 M 的著者。要关闭自动截词功能, 可将姓名加上双引号并限定在著者字段检索, 例如, 输入 "Smith M" [au], 可实现著者精确检索。

为提高著者检索的查准率, 可结合著者单位、主题等信息。例如, 输入 Smith MR Massachusetts General Hospital, 可以检索出哈佛大学麻省总医院 (Massachusetts General Hospital) Smith MR 的文献; 输入 Smith MR prostate cancer 可以检索出 Smith MR 有关前列腺癌方面的文献。此外, 2002 年以后的文献, PubMed 可实现对姓名全称的检索, 而且姓名排列顺序不限。

(2) 期刊检索: 可直接输入刊名全称、标准的 MEDLINE 刊名缩写或期刊的 ISSN 号。例如, 要查询《美国病理学杂志》的文献, 可输入刊名全称 "the American Journal of Pathology"、该刊的标准缩写 "am j pathol" 或者该刊的 ISSN 号 0002-9440。

当刊名与 MeSH 主题词相同时, 为避免误检, 可将刊名加双引号并限定在刊名字段, 例如, "Cell" [ta] 可检索出期刊 Cell 被 PubMed 收录的所有文献。

2. 高级检索 PubMed 的高级检索 (Advanced Search) 页面主要有检索输入框、检索构建器 (Search Builder) 及检索史 (Search History) 三部分 (图 5-2)。

(1) 检索构建器: 应用检索构建器可以很方便地实现多个字段的组合检索, 提高查准率; 结合检索历史的操作, 可完成复杂的布尔逻辑运算。

检索时, 先在左侧的下拉菜单中选择检索字段 (默认为 All fields), 输入检索词 (点击右侧的 "Show index list" 按钮, 系统显示该检索词的相关索引词, 可帮助正确选词), 选择布尔逻辑运算符 AND、OR 或 NOT, 上方的检索框中即显示输入的检索词及运算符, 点击 "Edit" 按钮可编辑检索式。完成检索式的构建后, 点击 "Search" 按钮, 返回检索结果。也可以点选 "Add to history" 按钮, 将检索式及其结果送入检索史中。

例如，检索哈佛大学麻省总医院 (Massachusetts General Hospital) Smith MR 发表的有关前列腺癌的随机对照临床试验的文献，用检索构建器的检索步骤如下：①在左侧的下拉菜单中点选 Author 字段，输入 Smith MR，点选 AND；②点选 Affiliation 字段，输入 Massachusetts General Hospital，点选 AND；③点选 Title/Abstract 字段，输入 prostat* cancer，点选 AND；④点选 Publication Type 字段，点击输入框右侧的 Show index list，下拉框中出现所有的“出版类型”选项及记录数，点选“randomized controlled trial”，完成检索式构建。页面上方的检索框中显示的检索式为：(((Smith MR [Author]) AND Massachusetts General Hospital [Affiliation]) AND prostat* cancer [Title/Abstract]) AND "randomized controlled trial" [Publication Type]，点击“Search”按钮 (图 5-2)，检索结果为 32 篇。

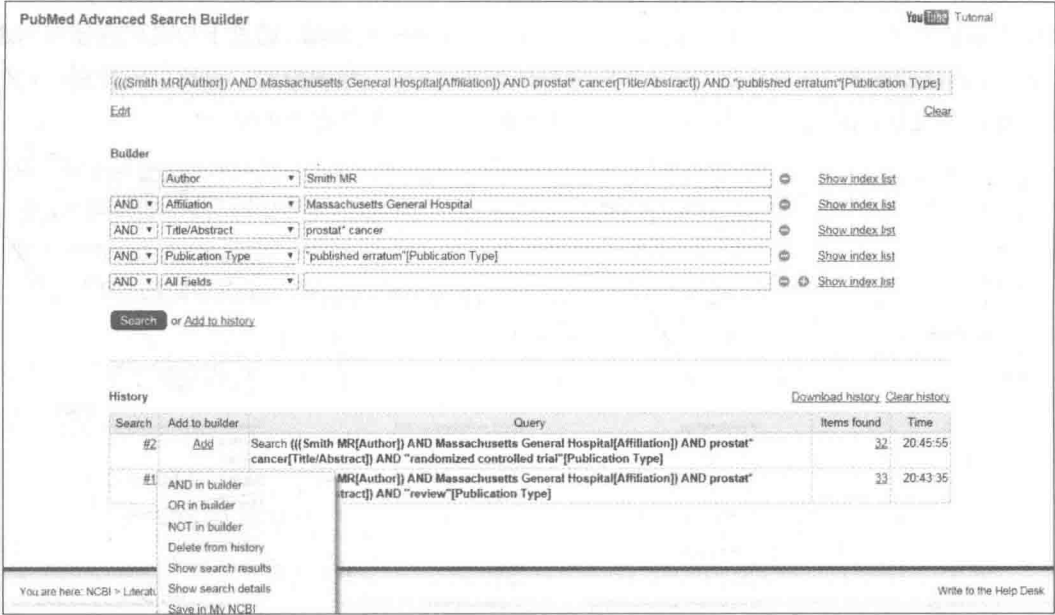


图 5-2 PubMed 高级检索页面

(2) 检索史：完整记录本次检索以来所有检索式的具体内容及命中结果，包括检索式序号、检索提问式 (Query)、检索结果数 (Items found) 及检索时间。单击检索式序号，可执行 AND、OR、NOT 逻辑运算，也可以删除检索式 (Delete from history)、显示检索结果 (Show search results)、显示实际执行的检索式 (Show search details) 或将检索式保存到 My NCBI (Save in My NCBI) (图 5-2)。检索历史最多保留 8 小时。此外，还可以点击“Add”按钮直接将某个检索式添加至检索构建器 (Add to builder)。

高级检索尤其适用于需要分步检索才能完成的复杂课题。实际检索时，在检索构建器中选择检索字段并输入检索词，点选“Add to history”按钮，将该检索词及检索结果送入到检索史；重复这一步骤，把所有检索词送入检索史后，再在检索框中直接输入用检索式序号构建的检索表达式，例如 (#3 OR #4) AND #5 NOT #7，再点击“Search”按钮完成检索。这样既简洁明了，又方便对检索表达式进行调整、修正。当然，也可以先在基本检索页面完成各个检索词的检索，再在检索史中构建完整的检索表达式。

3. 主题词检索 是 PubMed 最有特色的检索方法，能保证较好的查全率和查准率。在 PubMed 主页面或高级检索页面的“More Resources”栏目下点击 MeSH Database，进入 MeSH 主题词检索页面。有关医学主题词表的结构、内容及应用请参考本书第二章第五节。

主题词检索的步骤：①点击 MeSH Database 进入 MeSH 主题词检索页面；②输入检索词，点击“Search”按钮，页面显示与检索词相关的主题词列表，浏览选择合适的主题词；③点击选

定的主题词, 页面显示该主题词的详细信息, 包括主题词的定义、收录 MeSH 主题词表的年份 (Year introduced)、可匹配的副主题词 (Subheadings)、树状结构号 (Tree Number(s))、款目词 (Entry Terms, 也称入口词, 一般为主题词的同义词或相关词)、相关参照 (See Also)、历史注释 (Previous indexing) 等; ④选择副主题词: PubMed 共有副主题词 83 个, 每个副主题词均有其特定含义及使用范围, 主题词 / 副主题词的组配可使检索结果更专指。若无法确定合适的副主题词, 也可不选, 系统默认选择全部副主题词; ⑤其他检索限定: 点选“Restrict to MeSH Major Topic”表示将检索词限定在 Major Topic (主要主题词), 可提高查准率, 此外, 若点选“Do not include MeSH terms found below this term in the MeSH hierarchy”, 表示不扩展检索该主题词的下位词, 易造成漏检, 一般不建议点选; ⑥点击页面右上方 PubMed Search Builder (检索构建器) 的“Add to search builder”按钮, 其上的 PubMed Search Builder 中显示选定的主题词 / 副主题词及检索限定 (图 5-3); ⑦若检索课题有多个主题词, 可重复上述步骤, 通过下拉菜单选择布尔逻辑算符, 构建检索式, 最后点击“Search PubMed”按钮, 完成主题词检索。也可以先分别检索每个主题词的文献, 最后在高级检索的检索史中完成主题词的布尔逻辑组配。

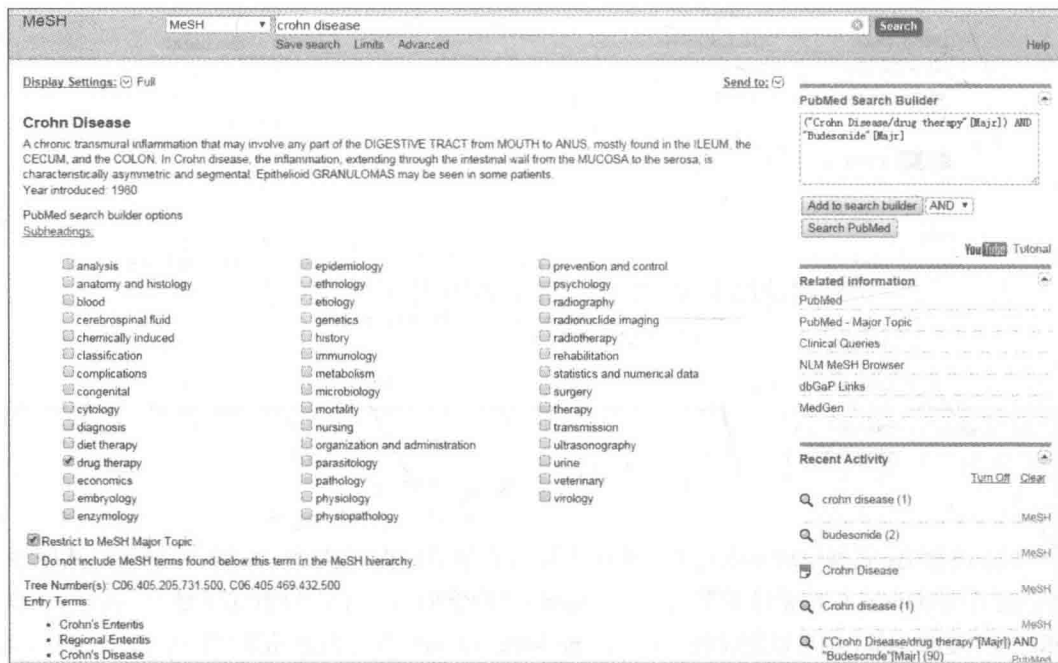


图 5-3 PubMed 主题词检索页面

例如, 要查找“糖皮质激素布地奈德 (budesonide) 治疗克罗恩病 (Crohn's disease)”的密切相关文献, 用主题词检索的基本过程如下: ①进入 MeSH 主题词检索页面, 输入 Crohn's disease, 点击“Search”按钮, 显示克罗恩病的 MeSH 主题词为 Crohn disease; ②点击该主题词, 选择副主题词 drug therapy (药物疗法), 勾选“Restrict to MeSH Major Topic PubMed”选项 (限定为主要主题词), 点击“Add to search builder”按钮; ③继续在该页面的检索框中输入 budesonide, 点击该主题词, 勾选“Restrict to MeSH Major Topic PubMed”, 点击“Add to search builder”按钮, 逻辑运算符选“AND”, 检索构建器中显示的检索式为 ("Crohn Disease/drug therapy" [Majr]) AND "Budesonide" [Majr], 点击“Search PubMed”按钮, 如图 5-3, 返回检索结果, 命中密切相关文献 90 篇。

主题词检索的优点很明显, 主题词对同一概念的不同表达方式进行了规范, 而且可以通过组配副主题词及限定主要主题词使检索结果更加专指、准确。同时, 利用主题词的树状结构表, 可以很方便地扩大或缩小检索范围。

需要注意的是：①主题词检索只针对 MEDLINE 数据库中的记录，不能检索其他来源的记录（包括 in-process citations, supplied by publisher 等），因此主题词检索可能漏掉一些已经被 PubMed 收录但尚未完成主题词标引加工的最新文献；② MeSH 主题词表每周更新，每年都会审核修订，但新出现的名词术语一般不会马上被主题词表收录；③ 2014 年 MeSH 词表中收录的主题词超过 27 000 个，而且可以通过 21.8 万个款目词及 21.9 万个补充概念（supplementary concept）自动转换为相应的主题词。尽管如此，仍有一些专指的医学概念没有直接对应的主题词。因此，实际检索中应根据课题情况，灵活应用各种检索方法，如主题词检索结合自动词语匹配检索，主题词检索结合多字段组合检索等，以达到满意的检索效果。

4. 临床查询（Clinical Queries）是专门为临床医生设计的检索服务。在 PubMed 主页面的“PubMed Tools”栏目下点击 Clinical Queries，即可进入临床查询页面，包括 Clinical Study Categories、Systematic Reviews 及 Medical Genetics 三类检索（图 5-4）。

（1）Clinical Study Categories：提供疾病的 etiology（病因）、diagnosis（诊断）、therapy（治疗）、prognosis（预后）及 clinical predictions guides（临床预测指南）等五个方面的查询，检索范围（Scope）可以选 Broad（敏感性检索，检索结果较宽泛）或 Narrow（特异性检索，检索结果较专指）。

（2）Systematic Reviews：专用于检索 systematic reviews（系统评价）、meta analysis（meta 分析）、reviews of clinical trials（临床试验综述）、guidelines（指南）等循证医学信息。

（3）Medical Genetics：专用于查询疾病的遗传学方面的文献，有 All（所有）、Differential Diagnosis（鉴别诊断）、Clinical Description（临床描述）、Management（处理）、Genetics Counseling（遗传咨询）、Molecular Genetics（分子遗传学）、Genetics Testing（遗传测试）等选项。

PubMed Clinical Queries

Results of searches on this page are limited to specific clinical research areas. For comprehensive searches, use PubMed directly.

Search:

Clinical Study Categories

Category:

Scope:

Results: 6 of 10213

Probiotics and probiotics in irritable bowel syndrome and inflammatory bowel disease in children.

Guandalini S, Cernat E, Moscoso D. *Benef Microbes*. 2014 Nov 12; 1-9. Epub 2014 Nov 12.

Vedolizumab for inflammatory bowel disease: Changing the game, or more of the same?

Raine T. *United European Gastroenterol J*. 2014 Oct; 2(5):333-44.

Vedolizumab for ulcerative colitis and Crohn's disease: results and implications of GEMINI studies.

Lam MC, Bressler B. *Immunotherapy*. 2014 Sep; 6(9):963-71.

[Chronic inflammatory bowel diseases. Strive for healing the mucosa, prevent overtreatment]

[No authors listed]. *MMW Fortsch Med*. 2014 Jul 24; 196(13):64-5.

Mesenchymal stromal cells from adipose tissue attached to suture material enhance the closure of enterocutaneous fistulas in a rat model.

Systematic Reviews

Results: 6 of 610

Systematic review. Histological remission in inflammatory bowel disease. Is 'complete' remission the new treatment paradigm? An IOIBD initiative.

Bryant RV, Winer S, Spi T, Riddell RH. *J Crohns Colitis*. 2014 Sep 26. Epub 2014 Sep 26.

Methotrexate for maintenance of remission in Crohn's disease.

Patel V, Wang Y, MacDonald JK, McDonald JW, Chande N. *Cochrane Database Syst Rev*. 2014 Aug 26; 8:CD006884. Epub 2014 Aug 26.

TNF inhibitors - Mechanisms of action, approved and off-label indications.

Cesak G, Kuzawirska O, Burda A, Lis K, Wojnar M, Mirowska-Guzel D, Balcowicz-Iskra E. *Pharmacol Rep*. 2014 Oct; 66(5):836-44. Epub 2014 May 29.

Budesonide for maintenance of remission in Crohn's disease.

Kuenzig ME, Rezaie A, Seow CH, Otley AR, Steinhart AH, Griffiths AM, Kappan GG, Benchimol EI. *Cochrane Database Syst Rev*. 2014 Aug 21; 8:CD002913. Epub 2014 Aug 21.

Medical Genetics

Topic:

Results

An inter... Genetic Counseling

Connelly TM, Koltun WA, Sangster W, Berg AS, Hegarty JP, Harns L 3rd, DeLong S, Stewart DB. *Surgery*. 2014 Oct; 156(4):769-74.

Crohn's disease in Korea: past, present, and future.

Lee KM, Lee JM. *Korean J Intern Med*. 2014 Sep; 29(5):558-70. Epub 2014 Aug 28.

A TNFSF15 disease-risk polymorphism increases pattern-recognition receptor-induced signaling through caspase-8-induced IL-1.

Hedi M, Abraham C. *Proc Natl Acad Sci U S A*. 2014 Sep 16; 111(37):13451-6. Epub 2014 Sep 2.

Osteopathic primary care of patients with inflammatory bowel disease: a review.

Bragg JD. *J Am Osteopath Assoc*. 2014 Sep; 114(9):695-701.

图 5-4 PubMed 临床查询页面检索

5. 单篇引文匹配器 单篇引文匹配器（Single Citation Matcher）主要用于查找某篇文献的准确信息。在 PubMed 主页面的“PubMed Tools”栏目下点击 Single Citation Matcher，即可在系统给出的选项中输入任何已知的信息，包括有效的刊名信息（全称或缩写均可）、出版年、卷、期、起始页码、作者、篇名中的任意词。点击“Search”按钮，系统返回符合要求的文献信息。

6. 期刊数据库检索 在 PubMed 主页面“More Resources”栏目下点击 Journals in NCBI Databases，进入 NLM Catalog（NLM 目录）的期刊数据库，可输入期刊的主题（Topic）、刊名全称、刊名缩写或 ISSN 号进行查询。期刊数据库提供了 PubMed 及 NCBI 其他数据库所收录期

刊的详细信息,包括刊名缩写、出版商、创刊年、出版国、出版频次、电子期刊网址、语种、ISSN号、被 MEDLINE 收录状态等。

三、检索结果的处理

(一) 检索结果过滤

PubMed 检索结果显示页面左栏提供了过滤(filters)选项,可以对检索结果进行精确的限定(图 5-5),主要包括:

(1) Article types (文献类型): 可限定感兴趣的文献类型,如 Clinical Trial (临床试验)、Review (综述),点击“More”可显示更多的文献类型。

(2) Text availability (文本可获得性): 可限定检索结果有 Abstracts (文摘)、Full text (全文链接)或 Free full text (免费的全文链接)。

(3) PubMed Commons (PubMed 空间): 选择是否显示读者在 PubMed 空间对文献的评论(Reader comments)。

(4) Publication dates (出版时间): 限定文献的出版年限,也可定制时间范围(Custom date range),按格式输入具体时间范围后,点选“Apply”(应用)按钮即可。

(5) Species (物种): 限定研究对象,可以选 Humans (人类)或 Other Animals (其他动物)。

(6) Subject (子集): 可限定检索结果为一些特定专题,包括 AIDS (艾滋病)、Bioethics (生物伦理学)、Cancer (癌症)、Complementary Medicine (补充医学)、Dietary Supplements (膳食补充剂)、History of Medicine (医学史)、Systematic Reviews (系统评价)、Toxicology (毒理学)、Veterinary Science (兽医学)等。

(7) Journal Categories (期刊类别): 可限定检索结果来自 Core clinical journals (临床核心期刊)、Dental journals (牙科学期刊)、MEDLINE (被 MEDLINE 收录的期刊)或 Nursing journals (护理学期刊)。

(8) Ages (年龄): 提供了从新生儿到老年人等 14 种不同类型的年龄选项。

(9) Search Fields (检索字段): 点击 Choose,在下拉框中选择要限定的检索字段,如著者单位、著者、题名、题名或文摘 (Title/Abstract)、刊名主题词 (MeSH Terms)、主要主题词 (MeSH

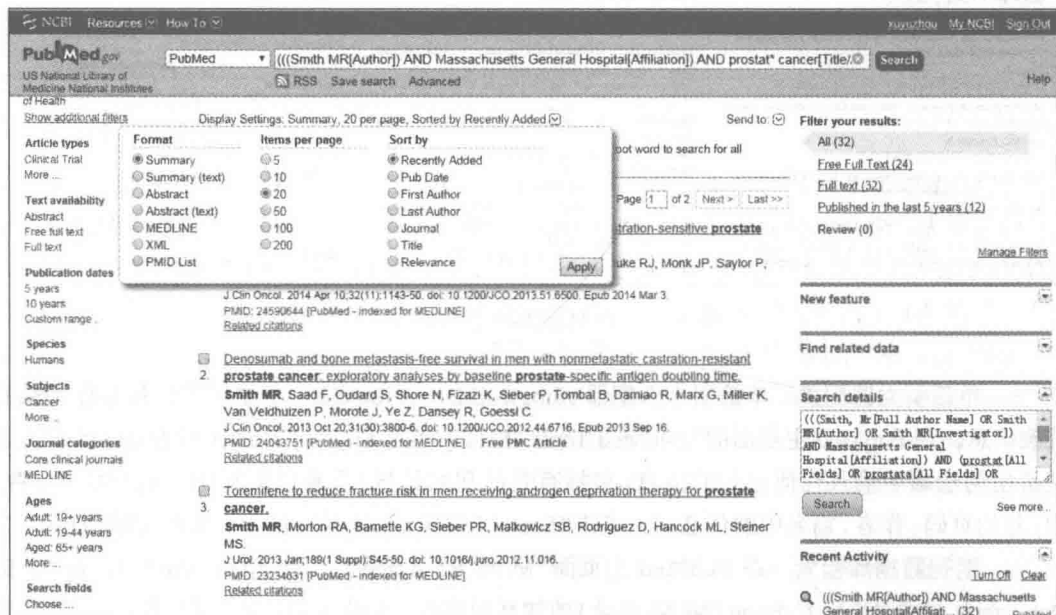


图 5-5 PubMed 检索结果页面 (Summary 格式)

Major Topic)等,点击“Apply”按钮,可将检索结果限定在某一特定字段。

此外,还可以对 Languages(语种)和 Sex(性别)进行限定。需要注意的是,选定了过滤条件后,系统会提示哪些选项被激活(“Filters activated: ...”),并在此后的检索中一直起作用。若要取消过滤条件,可点击“Clear all”或者在页面左栏直接取消已勾选的过滤条件。

(二) 检索结果显示

1. 显示格式 PubMed 的检索结果有多种显示格式,默认为 Summary 格式(简要格式,见图 5-5),包括每篇文献的篇名、著者、刊名、出版年月及卷期页码、DOI 号、电子出版日期、PMID 号、记录状态、相关文献(Related articles)链接,如果该文可以免费获取全文,则有 Free Article 或 Free PMC Article 链接。

点击文献篇名,进入 Abstract 格式(文摘格式,图 5-6),除了显示 Summary 格式的信息,还提供 Author information(著者单位信息)、Abstract(摘要)、Publications Types(出版类型)、MeSH Terms、Substances(物质)、Secondary Source ID(第二来源库 ID)、Grant Support(资助信息)、LinkOut - more resources(外部链接)、PubMed Commons(PubMed 空间)等信息。如果用户所在机构订购了该电子期刊全文,点击页面右上方的出版商图标链接即可获取全文。如果是在美国临床试验库注册登记的临床试验,页面右上方还链接到 ClinicalTrials.gov 的图标。此外,页面右侧还显示该文献的相关文献(Related citations in PubMed)、该文献在 PubMed Central 中的被引用信息(Cited by 4 PubMed Central articles)以及最近的操作记录(Recent Activity)。

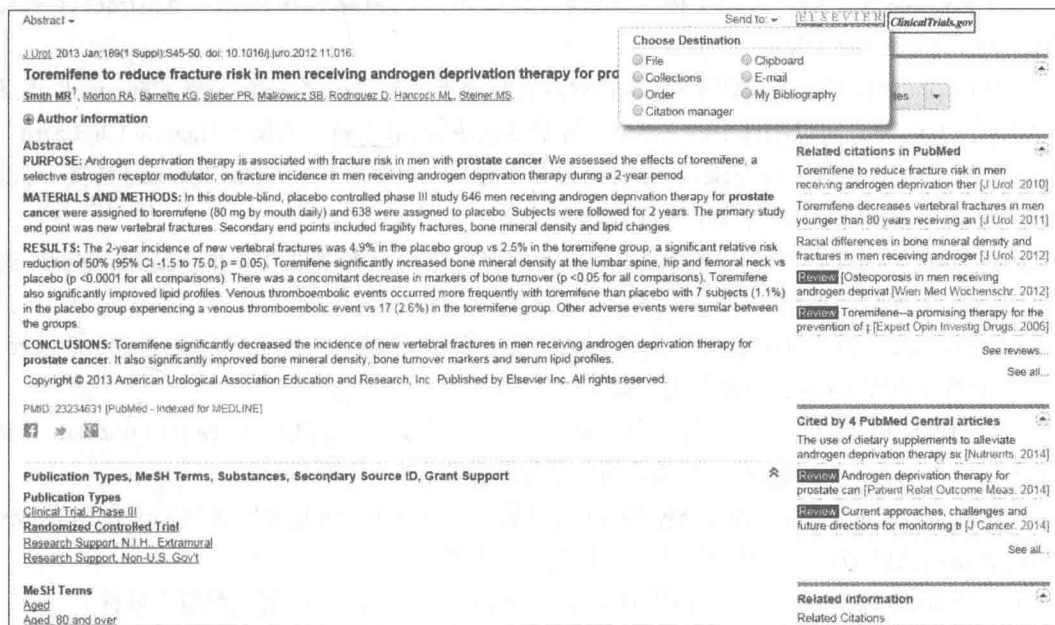


图 5-6 PubMed 检索结果页面(Abstract 格式)

除了上述两种基本格式外,点击检索结果页面上方的“Display Settings”(见图 5-5),可更改为其他显示格式(Format): ① Summary(text): 以纯文本形式显示 Summary 格式的所有信息; ② Abstract(text): 以纯文本形式显示 Abstract 格式的所有信息; ③ MEDLINE: 显示 MEDLINE 记录中的全部字段信息,并含有字段标识符; ④ XML: 显示 XML(可扩展标记语言)格式的记录信息,方便在 Web 环境下对数据进行转换与描述; ⑤ PMID List: 仅显示每条记录的 PMID 号。

2. 其他信息 PubMed 检索结果默认按 Recently Added(最近新增)排序,点击检索结果显示页面上方“Display Settings”的排序方式(Sort by),可以选择按 Pub Date(出版时间)、First Author(第一作者)、Last Author(排名最后的作者)、Journal(刊名)、Title(篇名)、Relevance(相关性)等排序,更改设置后点击“Apply”按钮(见图 5-5)。检索结果默认每页显示 20 条记录,点击

“Display Settings”的“Items per page”，可修改每页显示的记录条数。

检索结果显示页面右侧有多个栏目(见图 5-5)，主要包括：①检索结果过滤(Filter your results)：显示总的检索结果、综述文献、免费全文(Free Full Text)等信息，方便用户有选择地浏览感兴趣的检索结果。注册了 My NCBI 账号的用户可以点击“Manage Filters”链接，对检索结果过滤功能进行个性化的设置；②查找相关信息(Find related data)：可以选择在 NCBI 其他数据库(如 Books, Gene, Protein 等)中检索该课题，以扩大检索范围；③检索细览(Search details)：详细显示 PubMed 实际完成的检索式，特别是系统如何通过自动词语匹配功能将检索词转换为 MeSH 主题词，点击“See more”可以对 Query Translation 中的实际检索式进行修改，重新检索；④最近操作记录(Recent activity)：显示最近 8 小时内用户在 NCBI 的操作记录。其他栏目还包括 New Features(介绍 PubMed 的新特征)、Results by years(用直方图的形式显示每年的检索结果篇数)、Titles with your search terms(提供 20 篇标题中包含了检索词的文献)、Free full-text articles in PubMed Central(提供在 PubMed Central 中能获取免费全文的文献)以及 PMC Images search for(来自 PMC 的图片)等，但这些栏目不一定每次都在检索结果页面显示。

(三) 检索结果保存及输出

PubMed 提供了多种保存及输出检索结果的方式，点击检索结果显示页面的“Send to”下拉菜单，有 File、Clipboard(剪贴板)、Collections、E-mail、Order、My Bibliography、Citation manage 等多种检索结果保存及输出方式(图 5-6)。

(1) File：将选中的记录以文件形式保存，格式可以选 Summary(text)、Abstract(text)、XML、MEDLINE 或 PMID。

(2) Clipboard(剪贴板)：浏览检索结果时，在复选框中勾选所需文献，点选 Clipboard，再点击 Add to Clipboard，可以将选中的文献暂存在剪贴板中，同时文献下方显示“Item in Clipboard”。剪贴板中最多可保存 500 条记录，保存时间为 8 小时。点击页面右上方的 Clipboard 链接，可以显示所有暂存的文献记录，再集中进行处理。

(3) Collections：注册了 My NCBI 账号的用户可以将检索结果直接发送到 My NCBI 的 Collections，方便日后调用。

(4) E-mail：将选中的检索结果直接发送到指定的电子邮箱，一次最多可发送 200 条记录。选择显示格式及排序格式，再输入电子邮箱地址即可。

(5) Order：如果用户无法获取选中文献的全文，可点击 Order 通过 NLM 的 LoansomeDoc 文献传递系统订购全文。这是一项付费的文献传递服务。

(6) My Bibliography：注册了 My NCBI 账号的用户可以将检索结果发送到 My NCBI 的 My Bibliography(我的书目)，方便对检索结果进行后续管理。

(7) Citation manager：将检索结果直接导入 EndNote、NoteExpress 等文献管理软件。

此外，点击检索结果显示页面的 RSS 图标，再点击 Create RSS，即可订阅 RSS Feed，随时浏览追踪最新的检索结果。

四、个性化功能与服务

除了上述检索功能，PubMed 还有 My NCBI 个性化服务工具，LinkOut(外部链接)、Batch Citation Matcher(批量引文匹配器)、Clinical Trials(临床试验数据库)等链接，以及完善的帮助系统。

1. My NCBI 是一个功能强大且非常实用的文献管理工具。PubMed 主页的右上方有登录 My NCBI 按钮(见图 5-1)。用户注册并建立 My NCBI 账号后，通过 My NCBI 可以管理在 NCBI 所有数据库的检索过程及检索结果，追踪了解所在研究领域的最新进展，并实现个性化服务。My NCBI 的主要功能：



(1) 保存检索策略及定期发送最新结果: 点击结果显示页面检索框下方的“Save search”链接, 系统显示完整的检索策略 (Your PubMed search), 输入检索策略的名称 (Name of search), 点击“Save”按钮, 系统自动提示是否通过 Email 定期自动发送最新的检索结果, 可选择发送的频率、格式等。点击 My NCBI 的 Saved Searches 栏目之下的“Manage Saved Searches”链接, 可以对保存的检索策略进行删除、修改等 (图 5-7)。

(2) 保存检索结果: 在结果显示页面的“Send to”下拉菜单中选择 Collections, 点击“Add to Collections”按钮, 可输入文件名建立一个新文件保存检索结果, 也可以选择将检索结果添加到已有的文件记录中。点击 My NCBI 的 Collections 栏目“edit”链接可编辑已经保存的检索结果, 点击“Manage Collections”链接, 可对保存的检索结果进行删除、合并等操作 (见图 5-7)。

(3) 过滤检索结果: My NCBI 的过滤器 (Filters) 提供个性化的检索结果过滤功能。点击“Manage Filters”按钮, 进入过滤器管理页面: 左栏为已经激活的 PubMed 过滤器列表 (Your PubMed filter list), 右栏可浏览 / 检索并定制过滤器 (Browse/Search for PubMed Filters)。选择过滤器类别 (Select category), 有 Popular (通用)、LinkOut (外部链接)、Properties (特性)、Links (链接到 PubMed 其他数据库) 等 4 类, 例如可以在 Popular 中勾选 Free Full Text (免费全文)、Published in the last 5 years (最近 5 年的文献)、Review (综述), 选定后这些过滤项即被激活并添加到左栏的过滤器列表中。

(4) 个性化页面定制: 点击 NCBI Site Preferences, 可对 PubMed 的多项参数作个性化设置, 如输入检索词时自动提示 (Auto Suggest)、检索结果显示设定 (Result Display Settings)、检索词高亮 (Highlighting) 等。

此外, 利用 My NCBI 可以建立个人书目记录 (My Bibliography) 以及其他文献记录 (Other Citations)。Recent Activity 则可以显示最近 6 个月内用户的检索操作记录 (图 5-7)。

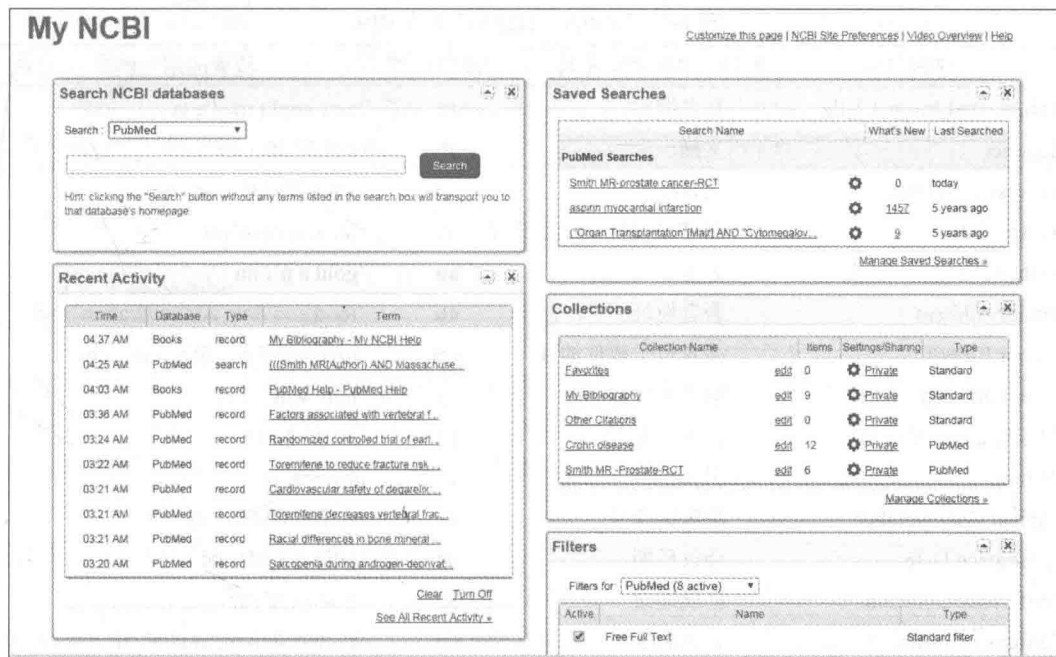


图 5-7 My NCBI 页面

2. LinkOut PubMed 主页“More Resources”栏目下有 LinkOut (外部链接) 按钮。PubMed 及 NCBI 的其他数据库通过 LinkOut 与互联网上的其他外部资源与服务建立链接, 包括在线全文数据库、图书馆馆藏信息、消费者健康信息、研究工具等。

3. 批量引文匹配器 点击 PubMed 主页的“PubMed Tools”栏目下的 Batch Citation Matcher,

进入批量引文匹配器，主要用于批量核对文献信息。检索时输入的格式为：刊名|年|卷|起始页码|著者|核对文献的标识，返回的结果含有文献记录的唯一识别码 PMID。用 Batch Citation Matcher 一次可核对不超过 100 条文献，检索结果可通过 Email 发送或直接存入文件中。

4. 临床试验数据库 点击 PubMed 主页“More Resources”栏目下的 Clinical Trials，进入 NLM 研制开发的临床试验数据库(<http://clinicaltrials.gov/>)。Clinical Trials 收录了美国及全世界 170 多个国家经注册登记的 8 万多个临床试验的详细信息，是十分重要的循证医学信息资源。

第二节 Embase

一、概 述

Embase 是生物医学和药学研究领域最重要的文摘数据库之一，最早源自 1947 年荷兰 Elsevier (爱思唯尔) 公司出版的印刷型检索工具 *Excerpta Medica* (EM,《医学文摘》)，20 世纪 70 年代发展为 EM 文摘数据库。Embase 整合了 EM 文摘数据库与 MEDLINE 数据库的全部内容，并去除了重复记录。

Embase 收录了全世界范围内 8400 多种生物医学和药学方面的同行评议期刊，数据每个工作日更新，每年新增文献记录 130 多万条，总文献记录超过 2800 万条，其中包括 MEDLINE 数据库未收录的 2700 多种期刊以及 600 多万条文献记录，文献可回溯至 20 世纪 40 年代早期。Embase 还收录了 2009 年以来重要的生物医学、药物以及医疗设备方面的会议录信息，每年收录来自 1000 多个会议的 30 万条会议录文摘。

Embase 的记录字段共有 43 个，所有字段均为可检索字段，字段名称、字段标识符见表 5-2。

表 5-2 Embase 数据库字段结构表

字段名称	字段中文译名	字段标识符	检索应用范例
Abbreviated Journal Title	刊名缩写	ta	'new engl j med': ta
Abstract	文摘	ab	'end of life care': ab
Accession Number	文献编号	an	'2008579230': an
Article Title	题名	ti	'bioterrorism': ti
Author Name	著者	au	'gold a p': au
Author Address	著者地址	ad	'food and drug administration': ad
Author E-mail	著者电子邮件地址	em	'myname@goodisp.com': em
Book Publisher	图书出版商	bp	'john wiley': bp
CAS Registry Number	CAS 登记号	rn	'50 99 7': rn
Citation	引文	ct	physics: ct
Clinical Trial Number	临床试验号	cn	NCT00079001: cn
Conference Date	会议日期	dc	'1989-05-04': dc
Conference Location	会议地点	lc	'new york': lc
Conference Name	会议名称	nc	'symposium on niddm': nc
Country of Author	著者所在国	ca	italy: ca
Country of Journal	期刊出版国	cy	canada: cy
Device Manufacturer	设备厂商	df	dupont: df
Device Trade Name	设备商标名称	dn	dermabond: dn
Drug Index Term	药物标引词	dd	diazepam: dd; simvastatin/dd
Drug Manufacturer	药物制造商	mn	glaxo: mn

续表

字段名称	字段中文译名	字段标识符	检索应用范例
Drug Trade Name	药物商标名称	tn	lipitor: tn
Editor Name	编辑者	ed	johnson: ed
EMBASE Classification	纸本 EM 分册号	cl	'047': cl
Index Term (Emtree Subject Descriptor)	Emtree 主题词	de	'insulin': de
Explosion (Emtree Index Term)	扩展词	exp	'antineoplastic agent'/exp
ISBN	国际标准书号	ib	'9781573317351': ib
ISSN	国际标准刊号	is	'0012-186x': is
Issue	期号	ip	'10': ip
Journal Title	刊名	jt	'new england journal of medicine': jt
Language of Article	原文语种	la	french: la
Language of Summary	文摘语种	ls	english: ls
Link (Free-floating Subheading)	副主题词	lnk	'adverse drug reaction': lnk
Major (Emtree index term)	主要主题词	mj	Ischemia/mj
Medical Index Term	医学标引词	dm	'lung tuberculosis': dm; heart/dm
Molecular Sequence Number	分子序列号	ms	eu518933: ms
Original (Non-English) Title	非英文原文题名	tt	(herz or coeur): tt
Page Range	起止页码	pg	'61-88': pg
Publication Type	出版类型	it	'short survey': it
Publication Year	出版年	py	'2009': py; [2007-2009]/py
Publisher Item Identifier	出版商标识号	ii	S0093691x02011287: ii
Source Type	资源类型	pt	book: pt
Start Page	起始页码	sp	'27': sp
Volume	卷号	vi	'39': vi

二、检索方法与技巧

(一) 检索规则

- 1. 布尔逻辑运算符 支持 AND、OR 和 NOT 三种布尔逻辑运算。
- 2. 邻近算符 支持“NEAR/n”和“NEXT/n”两种邻近算符，两者均表示连接的两个检索词之间相隔不能超过 n 个单词，但“NEAR/n”对两词的前后顺序没有要求，“NEXT/n”则要求两词的前后顺序不能改变。
- 3. 截词符 支持“*”和“?”两种截词符，其中“*”号表示零个或多个字符，“?”号表示 1 个字符，两种截词符均可置于单词词尾或词间。
- 4. 短语检索 将短语加上引号表示精确查找某一短语或词组，此时系统不再自动对词组进行拆分。含有连字符“-”的短语，系统也不进行拆分。短语检索不支持邻近算符和截词符。
- 5. 字段限定符 有“:”和“/”两种。字段限定符“:”可用于所有字段，并可同时限定多个字段，字段标识符之间用逗号分隔；字段限定符“/”仅用于对部分字段进行精确限定检索，即实现与检索词完全一致的检索。表 5-2 给出了字段限定检索的应用范例。

(二) 检索方法

Embase 检索功能强大，提供快速检索、高级检索、药物检索、疾病检索、设备检索、文章检索、主题词检索、期刊浏览和著者检索等多种检索途径。

1. 快速检索(Quick Search) 进入 Embase 主页,页面上方的工具条包括 Search, Browse, Results, Tools 四个部分,默认为快速检索页面(图 5-8)。在检索框中直接输入单词、短语词组(要求用引号括起)或检索表达式,系统在 Embase 全字段范围内检索。输入检索词时,系统自动启动“autocomplete”功能,利用 Emtree 主题词表指引并提示用户可以选择哪些合适的主题词进行检索。

图 5-8 Embase 主页面(快速检索)

快速检索时,默认勾选“search as broadly as possible (map, explode, search also as free text in all fields)”选项(见图 5-8),即尽可能执行宽泛的检索,系统会自动在 Emtree 主题词表中对检索词相对应的主题词及其下位词进行检索,同时将该检索词作为自由词在全字段检索,再执行“OR”逻辑运算,这个过程类似 PubMed 的自动词语匹配检索。快速检索页面还提供了检索年限的选项。

2. 高级检索(Advanced Search) 在 Embase 主页 Search 下拉菜单中选择“Advanced”,进入高级检索页面(图 5-9)。高级检索的检索规则及检索表达式的编制方法与快速检索相同,但高级检索提供了多种对检索式进行修饰和限定的选项,检索功能更加强大,可以实现精细检索以及对复杂课题的检索。

(1) Mapping: Embase 默认执行与快速检索相同的自动词语匹配检索,即尽可能执行宽泛的检索(Search as broadly as possible),将检索词自动匹配为 Emtree 中相对应的主题词(Map to preferred term in Emtree),并对主题词的下位词进行扩检(Explode using narrower Emtree terms),同时将检索词作为自由词在全字段检索(Search also as free text in all fields)。如果勾选 Limit to terms indexed in article as ‘major focus’ 选项,表示在主题词匹配检索时为仅检索主要主题词(图 5-9)。

图 5-9 Embase 高级检索页面

(2) Fields: 列出了 Embase 的 37 个常用字段及标识符, 可以将检索词限定在某些特定字段。

(3) Quick limits (快速限定): 提供 Humans, Animals, With abstracts, Priority journals (Embase 选定的 1800 多种优先期刊), Only in English, Article in Press (出版中的文献), in Process (处理中的文献), With molecular sequence number (有分子序列号), With clinical trial number (有临床试验注册号) 等 9 类快速限定选项。

(4) EBM (循证医学): 可限定检索结果为 Cochrane Review (Cochrane 系统评价)、Systematic Review (系统评价)、Controlled Clinical Trials (临床对照试验)、Randomized Controlled Trials (随机对照试验)、Meta Analysis (meta 分析) 等不同类型的循证医学文献。

(5) Pub.types (出版类型): 可限定检索 Article (论文)、Article in Press (出版中的论文)、Conference Paper (会议论文)、Review (综述) 等不同出版类型的文献。

(6) Age (年龄): 对涉及人的研究, Embase 提供了从胚胎、胎儿、新生儿直到老年 (65 岁以上)、高龄老人 (80 岁以上) 等 13 个不同年龄段的检索限定选项。

(7) Animals (动物): 可限定检索 Animal Cell (动物细胞)、Animal Experiment (动物试验)、Animal Model (动物模型)、Animal Tissue (动物组织) 等不同类型的实验研究文献。

此外, 高级检索还提供 Date (日期)、Sources (来源, 即限定仅检索 Embase 或 MEDLINE)、Language (语种)、Gender (性别) 等限定选项。

例如, 要查找有关青少年糖尿病的综述文献, 可以在高级检索的检索框中输入 diabetes, 点击下方检索限定栏的“Age”, 勾选“Adolescent: 13 to 17 years”(青少年: 13~17 岁) 选项, 再点击检索限定栏的“Pub.types”, 勾选“Review”选项, 最后点击“Search”按钮。

3. 药物检索 (Drug Search) 药物检索是 Embase 的特色检索途径。在 Embase 主页上方 Search 下拉菜单中选择“Drug”, 进入药物检索页面 (图 5-10)。检索时, 可输入药物的通用名 (generic name)、商品名 (trade name)、实验室代码 (laboratory code) 或化学名 (chemical name), 系统自动将检索词强制转换为相匹配的 Emtree 药物主题词进行检索。例如, 输入抗溃疡药物奥美拉唑的商品名 Losec (洛赛克), 系统自动提示 losec use: omeprazole, 并将其转为药物主题词 Omeprazole。此外, 药物检索时, 检索结果显示页面的右侧会提供从 Elsevier 的药物安全信

Drug Search

e.g. low molecular weight heparin

Search Mapping Date Drug fields Sources Drug subheadings Routes Quick limits EBM Pub. types Languages

Routes of drug administration Clear page selections Collapse

☐ Buccal drug administration	☐ Intrabursal drug administration	☐ Intracisternal drug administration
☐ Epidural drug administration	☐ Intracameral drug administration	☐ Intradermal drug administration
☐ Inhalational drug administration	☐ Intracardiac drug administration	☐ Intraduodenal drug administration
☐ Intraarterial drug administration	☐ Intracavernous drug administration	☐ Intragastric drug administration
☐ Intraarticular drug administration	☐ Intracerebral drug administration	☐ Intralesional drug administration
☐ Intrabronchial drug administration	☐ Intracerebroventricular drug administration	☐ Intralymphatic drug administration
☐ Intramuscular drug administration	☐ Intraspinal drug administration	☐ Intrauterine drug administration
☐ Intranasal drug administration	☐ Intrathecal drug administration	☐ Intravaginal drug administration
☐ Intraocular drug administration	☐ Intratracheal drug administration	☐ Intravenous drug administration
☐ Intraosseous drug administration	☐ Intratumoral drug administration	☐ Intravesical drug administration
☐ Intraperitoneal drug administration	☐ Intratympanic drug administration	☐ Intravitreal drug administration
☐ Intrapleural drug administration	☐ Intraurethral drug administration	☐ Oral drug administration
☐ Parenteral drug administration	☐ Subcutaneous drug administration	
☐ Periocular drug administration	☐ Sublabial drug administration	
☐ Rectal drug administration	☐ Sublingual drug administration	
☐ Regional Perfusion	☐ Topical drug administration	
☐ Retrobulbar drug administration	☐ Transdermal drug administration	
☐ Subconjunctival drug administration		

☐ OR ☐ AND

图 5-10 Embase 药物检索页面

息数据库 PharmaPendium 获取的药物基本信息,包括药物名称、化学结构、分类、作用靶点、适应证、CAS 号以及药物副作用等。

药物检索的检索式修饰与限定方法与高级检索基本相同,只是针对药物检索的需要而增设了药物副主题词(Drug subheadings)与给药途径(Routes)两种检索限定选项,以增强检索深度,提高查准率。

Embase 有 17 个药物副主题词,包括 Adverse drug reaction(药物不良反应)、Clinical trial(临床试验)、Drug analysis(药物分析)、Drug combination(药物配伍)、Drug comparison(药物对照)、Drug concentration(药物浓度)、Drug development(药物开发)、Drug dose(药物剂量)、Drug interaction(药物相互作用)、Drug therapy(药物治疗)、Drug toxicity(药物毒性)、Endogenous compound(内源性物质)、Pharmaceutics(药剂学)、Pharmacoeconomics(药物经济学)、Pharmacokinetics(药代动力学)、Pharmacology(药理学)。使用药物副主题词可以使检索结果更专指,更准确。此外,Embase 提供了 47 种给药途径选项(见图 5-10),可以对给药方式进行深入细致的限定,如 Sublingual 舌下给药、Inhalational 吸入给药、Intracerebral 脑内给药、Regional perfusion 局部灌注等。

实际检索中,可以将多个检索限定选项组合使用,以提高查准率。例如,要查找化疗药物顺铂(cisplatin)动脉内给药的毒副作用的文献,首先进入 Drug Search 页面,在检索框中输入 cisplatin,点击下方检索限定栏的“Drug subheadings”,勾选药物副主题词“Adverse drug reaction”与“Drug toxicity”(默认两个副主题词之间的逻辑关系为“OR”),再点击检索限定栏的“Routes”,勾选“Intraarterial drug administration”(动脉内给药)选项,最后点击“Search”按钮。

4. 疾病检索(Disease Search) 疾病检索是系统专为检索疾病信息而设计的检索途径,点击 Embase 主页上方 Search 下拉菜单中的“Disease”,进入疾病检索页面。疾病检索时,系统会自动将疾病名称检索词强制转换为相匹配的 Emtree 疾病主题词进行检索。疾病检索的检索式修饰和限定方法与高级检索基本相同,只是针对疾病检索的需要而增设了 14 种疾病副主题词(Disease subheadings),分别是 Complication(并发症)、Congenital disorder(先天性疾病)、Diagnosis(诊断)、Disease management(疾病处理)、Drug resistance(药物抵抗)、Drug therapy(药物治疗)、Epidemiology(流行病学)、Etiology(病因学)、Prevention(预防)、Radiotherapy(放射治疗)、Rehabilitation(康复)、Side effect(副作用)、Surgery(手术)、Therapy(治疗)等。

5. 设备检索(Device Search) 是 Embase 针对医疗器械、设备提供的特色检索途径。点击 Embase 主页上方 Search 下拉菜单中的“Device”进入设备检索。检索方法及检索限定的内容都与高级检索类似,并增设了 Adverse device effect(设备副作用)、Device comparison(设备比较)、Device economics(设备经济学)与 Clinical trial(临床试验)等 4 个设备副主题词(Device subheadings)。

6. 文章检索(Article Search) 点击 Embase 主页上方 Search 下拉菜单中的“Article”即可进入文章检索。文章检索一般通过利用某些已知信息来准确查找、定位某篇或某类特定文献。文章检索提供标题、作者、期刊名称、刊名缩写、ISSN 号码、期刊卷、期号及文章起始页等多个字段的组合检索,还可限定出版日期。

7. Emtree 主题词检索 Emtree 是 Embase 对生物医学和药学文献进行主题分析、标引和检索时使用的权威主题词表。截至 2014 年 9 月,Emtree 有主题词超过 7 万个(其中药物和化学物质名词超过 3.1 万个),同义词超过 29 万个(其中药物和化学物质名词超过 18 万个),不但涵盖了 MeSH 词表的所有主题词,其数量更是 MeSH 主题词的两倍多。Emtree 还收录了 3000 多个医疗器械(设备)专有名词,数千个相关的医疗程序(medical procedures)术语(如 endoscopy,内镜检查术),64 个药物相关的副主题词(其中 47 个为给药途径),4 个医疗设备副主题词,14 个疾病副主题词,38 个研究类型标记词(38 check tags)。Emtree 与 2.3 万个 CAS 化学物质登记号



建立了链接,每年更新 3 次,以保证及时纳入在 WHO 登记的最新国际非专利药物(international non-proprietary names, INNs)以及美国 FDA 和欧洲药学会(European Medicine Agency)列出的所有新批准药物。此外, Emtree 也有类似 MeSH 主题词的树状结构表,所有主题词按学科领域共分为 14 个大类,从一般到专指,层层划分,形成树形结构的等级体系(tree hierarchy),可以很方便地扩大或缩小主题词检索的范围。

借助 Emtree 大容量的主题词表及强大的同义词库, Embase 对每条文献记录基于全文进行主题词的深度标引。因此,在主题词检索时,从药物、疾病、医学设备及生物医学名词术语的任一相关同义词入手,都能自动匹配为相对应的主题词,大大提高了检索效率,保证了较好的查全率和查准率。

在 Embase 主页上方 Browse 下拉菜单中选择“Emtree”,即可进入主题词检索页面(Browse Emtree)。系统提供查找和浏览两种方式。

(1) 查找主题词(Find Term): 在检索框内输入检索词或者词组,点击 Find Term 按钮,系统将按字顺轮排方式显示所有包含该词的同义词(款目词)和主题词。同义词用“use:”指引到相应的主题词(图 5-11)。点击主题词,显示该主题词在 Emtree 树状结构表中的位置,并直接显示标引有该主题词的文献篇数(点击该链接可直接浏览这些文献)。页面下方为该主题词的历史注释(History)、同义词(Synonyms)以及道兰医学辞典里与该词有关的注释信息。

系统默认勾选“Explosion”选项,对下位词进行扩展检索(Extent your search)。若勾选“As major focus”选项,表示限定在主要主题词检索。点击“Add to Query Builder”按钮可以将选中的主题词送入检索式编制输入框(Query Builder),点击“Search”按钮完成主题词检索。也可以将主题词送入高级检索、药物检索、疾病检索或设备检索,进一步组配副主题词、进行各种检索限定等(见图 5-11)。对涉及多个主题词的课题,可重复以上步骤,在 Query Builder 完成复杂的逻辑组配检索。

(2) 按类浏览主题词(Browse by Facet): Embase 提供按等级分类浏览主题词。点击“Browse by Facet”按钮,系统显示 Emtree 主题词表的 14 个大类,点击任意需浏览的类名,显示该类的下位类主题词,可层层点击逐级浏览,直到最底层的下位类。

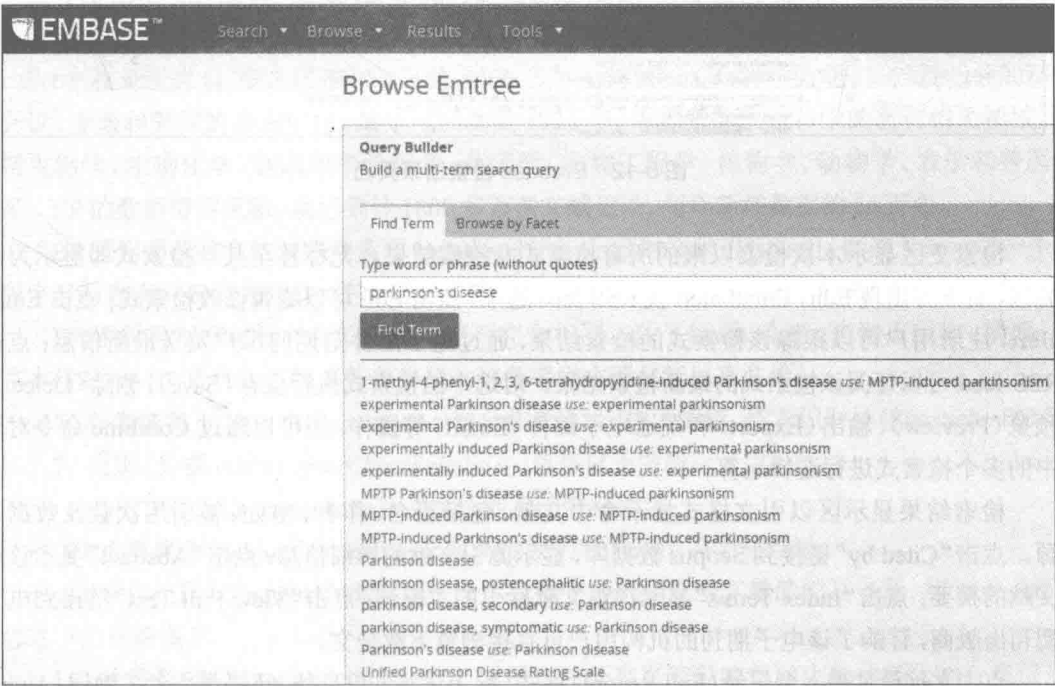


图 5-11 Emtree 主题词检索页面

8. 期刊浏览 (Journals) 点击 Embase 主页上方 Browse 下拉菜单的“Journals”, 进入期刊浏览 (Browse Journals) 页面。系统提供按刊名字顺浏览 Embase 收录的期刊 (不包括 MEDLINE 独有的期刊), 点击刊名右侧的“about”链接, 显示期刊的基本信息, 包括刊名全称、出版商、出版国以及出版频率。可逐级点击刊名、卷、期, 直到浏览某一期的文章。

9. 著者检索 (Authors) 点击 Embase 主页上方 Browse 下拉菜单的“Authors”, 进入著者检索 (Find Author) 页面。检索时可输入作者姓名的一部分, 姓在前用全称, 名在后用首字母, 姓与名之间空一格, 名与名之间用“.”或空格分隔。系统将显示所有前方一致的著者姓名。

例如, 要检索中山大学药理学的颜光美教授被 Embase 收录的文献, 首先在 Find Author 页面输入 Yan G, 系统显示所有 Yan G 开头的著者, 为防漏检, 浏览选择 Yan G.、Yan G.-M.、Yan G.-M.、Yan G.M.、Yan Guang-Mei 等各种表达方式, 点击“Find articles by these authors”按钮, 返回检索结果; 第二步在检索框中输入检索式: (Sun Yat-Sen AND Pharmacology): ad, 点击“Search”按钮, 系统检索出著者地址中含有 Sun Yat-Sen 及 Pharmacology 的所有文献; 最后勾选这两次检索的检索式序号左侧的复选框, 在检索史区的工具栏中点击“Combine using And”按钮, 完成检索, 显示检索结果为 84 篇, 均为符合要求的文献记录。

三、检索结果的处理

Embase 检索结果页面主要分为检索史区和检索结果显示区两个部分 (图 5-12)。

图 5-12 Embase 检索结果页面

检索史区显示本次检索以来的所有检索式及检索结果。光标移至某一检索式即显示为高亮区, 右下方出现 Edit, Email alert 及 RSS feed 选项: 点击 Edit 可以编辑修改检索式; 点击 Email alert, 注册用户可以跟踪该检索式的检索结果, 通过电子邮件定期向用户发送最新信息; 点击 RSS feed 可以订阅该检索式的最新检索结果。对选中的检索式执行保存 (Save)、删除 (Delete)、预览 (Preview)、输出 (Export) 和发送电子邮件 (Email) 等操作; 还可以通过 Combine 命令对选中的多个检索式进行逻辑运算。

检索结果显示区以引文格式显示命中文献, 包括篇名、作者、出处、被引用次数及数据来源。点击“Cited by”链接到 Scopus 数据库, 显示施引文献的详细信息; 点击“Abstract”显示该篇文献的摘要; 点击“Index Terms”显示该篇文献标引的主题词; 点击“View Full Text”链接到电子期刊出版商, 订购了该电子期刊的机构用户可直接浏览下载全文。

点击文献篇名进入全记录 (Full Record) 格式, 除上述基本信息外, 还提供作者关键词 (Author Keywords), 通讯作者地址 (Correspondence Address), 其他信息 (Additional Information), 包括刊



名缩写、ISSN 号、文献出版日期与入库日期、出版类型、语种、Embase 索取号、参考文献数目、CAS 登记号、药物制造商等。

检索结果默认按出版日期 (Publication) 排序, 也可以按相关度 (Relevance) 或者出版年 (Publication Year) 排序 (Sort by)。

检索结果显示区的左侧为检索结果过滤器 (Results Filters), 可对检索结果作进一步的提炼或筛选。过滤器包括来源、药物、疾病、设备、浮动副主题词、年龄、性别、研究类型、发表类型、作者、会议摘要、药物商品名、药物制造商、设备商品名、设备制造商等 16 种。实际应用时, 点击并展开感兴趣的过滤器, 选择相应的过滤内容, 点击“Apply”按钮, 即可显示过滤后的检索结果。

Embase 提供显示 (View)、打印 (Print)、输出至参考文献管理软件 (Export)、Email、原文订购 (Order)、添加至剪贴板 (Add to Clipboard) 等多种输出方式 (见图 5-12)。对浏览选中的文献, 点击“Add to Clipboard”可将文献暂存至剪贴板, 选中所有文献后, 再点击主页上方 Tools 下拉菜单中的 Clipboard, 将剪贴板中暂存的文献批量输出。

四、个性化功能与服务

Embase 提供个性化服务功能。点击主页右上角的 Register 链接, 可在系统中免费注册个人账户。登录账号后, 用户可将检索策略及检索结果保存在个人文档中, 在随后的检索中, 可对存储的检索策略进行编辑、修改和重新检索, 实现定题信息服务; 也可以通过个人账户实现 Email Alert 功能。

第三节 BIOSIS Previews

一、概 述

BIOSIS Previews (BP) 是世界知名的有关生命科学研究的文摘数据库, 由美国生物科学信息服务社 (BIOSIS) 编辑出版。BP 目前主要整合在 Web of Science (WOS) 和 Ovid 检索平台上, 本节以 WOS 平台上的 BP 为例, 介绍其使用方法。

BP 收录了来自 90 个国家和地区的 6000 多种生物学和生命科学的期刊, 以及相关的国际会议、专著和美国的专利信息, 最早可回溯至 1926 年, 内容覆盖生命科学的所有相关领域, 包括生物学、生物化学、临床和实验医学、药理学、生物工程学、植物学、动物学、农学和兽医学等。BP 的数据每周更新, 现已累计 1800 多万条文献记录, 每年新增数据约 56 万条。

BP 采用系统独有的关联性索引 (relational indexing) 对文献进行标引, 能深入揭示每一个检索字段与索引词表的关联性, 使用户能从多个字段迅速准确地找到相关文献。

BP 不同类型文献所包含的字段可能不完全相同。除了标题、作者、来源出版物、摘要、等基本信息外, BP 还标引了很多特色字段。这些字段主要包括以下几类:

1. 作者信息 (Author Information) 包括作者地址、电子邮箱、作者识别号 (ResearcherID) 等;
2. 类别 / 分类 (Categories/Classification) 包括以下字段:
 - (1) 研究方向 (Research Area): 文献所涉及的学科领域;
 - (2) 主要概念 (Major Concepts): 文献所涉及的较大范围的学科领域;
 - (3) 概念代码 (Concept Code): 用一个五位数字的代码反映文献的学科主题, 相当于主要概念之下的次级概念;
 - (4) 分类数据 (Taxonomic Data): BP 采用自然分类系统反映每种生物体的生物分类信息。在自然分类系统中, 最上位的类目分为微生物界、植物界、古生物界及动物界等 4 个大类, 大类之

下再分为门、纲、目、科等,均采用拉丁学名。分类数据包括:① SUPER TAXA:生物分类中较高级别的生物分类拉丁学名,一般按照从低级分类到高级分类显示;②分类注释(Taxa Notes):上位生物分类的俗名;③生物体分类[生物物种分类代码](Organism Classifier):上位生物分类之下更具体的生物分类拉丁学名及相应的生物物种分类代码;④生物体名称(Organism Name);⑤不同形式(Variant):生物物种名称的不同形式、其他常见名称或物种名称;⑥详细信息(Details):生物体的其他更为详细的信息,如生物的性别、发育阶段和作用。

(5) 疾病数据(Disease Data):文献涉及的疾病信息,用 MeSH 主题词标引,包括疾病名词、MeSH 主题词、疾病类型(Disease Affiliation, 疾病上位词)、疾病详细信息(一般为副主题词)。

(6) 化学数据(Chemical Data):文献涉及的化学和生化物质(包括药物)信息,一篇文献最多标引 20 种化学和生化物质,包括化学名称、化学名称的不同形式、化学物质 CAS 登记号(CAS Registry No.)、药品限定词(Drug Modifier, 用于定义文献中所涉及的化学物质的治疗作用,一般是比较宽泛的概念)、酶学委员会(化学物质的酶学委员会编号)、化学详细信息等。

(7) 基因名称数据(Gene Name Data):文献涉及的基因信息,包括基因名称、不同形式、详细信息等。

(8) 综合叙词(Miscellaneous Descriptors):BP 的索引系统将不能纳入上述字段的其他主题词自动分配到综合叙词字段。

3. 文献信息(Document Information) 包括出版类型、媒介、文献类型、语种、入藏号(Accession Number, BP 中每条记录的唯一索取号)、PubMed ID、ISSN 号及 eISSN 号。

4. 其他信息(Other Information) 包括以下字段:

(1) 方法和设备(Methods and Equipment Data):文献涉及的方法和设备的名称、不同形式、详细信息等。

(2) 器官/系统/细胞器数据(Parts and Structures Data):文献涉及的器官/系统和细胞器数据。

(3) 地理数据(Geographic Data):文献涉及的地理数据信息。

5. 会议(Conference) 包括会议名称、会议地点、会议日期及会议主办者等信息。

此外,对专利文献,BP 还提供发明人(Inventor(s))、专利号(Patent Number)、专利授权日期(Patent Date Granted)、专利国/地区(Patent Country)、专利权人(Patent Assignee)、专利分类号(Patent Class)等信息。

二、检索方法与技巧

(一) 检索规则

整合在 WOS 平台的 BP 采用与 WOS 平台数据库相同的检索规则,运算符和检索规则如下:

1. 支持布尔逻辑算符 AND、OR、NOT,运算符不区分大小写。
2. 支持“*” (代表 0 到多个字母)、“\$” (代表 0 到 1 个字母)及“?” (代表 1 个字符)三种截词符。
3. 支持 NEAR/x 及 SAME 两种位置算符。其中 NEAR/x 表示 NEAR 连接的两个检索词之间最多相隔不超过 x 个单词。SAME 仅用于地址字段,表示 SAME 连接的检索词出现在同一地址中。
4. 支持短语精确检索,将短语加上双引号。
5. 运算符的优先顺序为 NEAR/x > SAME > NOT > AND > OR,使用括号可改变优先顺序,括号内的检索式优先执行。

(二) 检索方法

1. 基本检索 进入 BP 主页面(图 5-13),默认为基本检索,可转至高级检索,也可切换到



WOS 平台的其他数据库。检索区下方的时间跨度选项可提供检索年限限定。

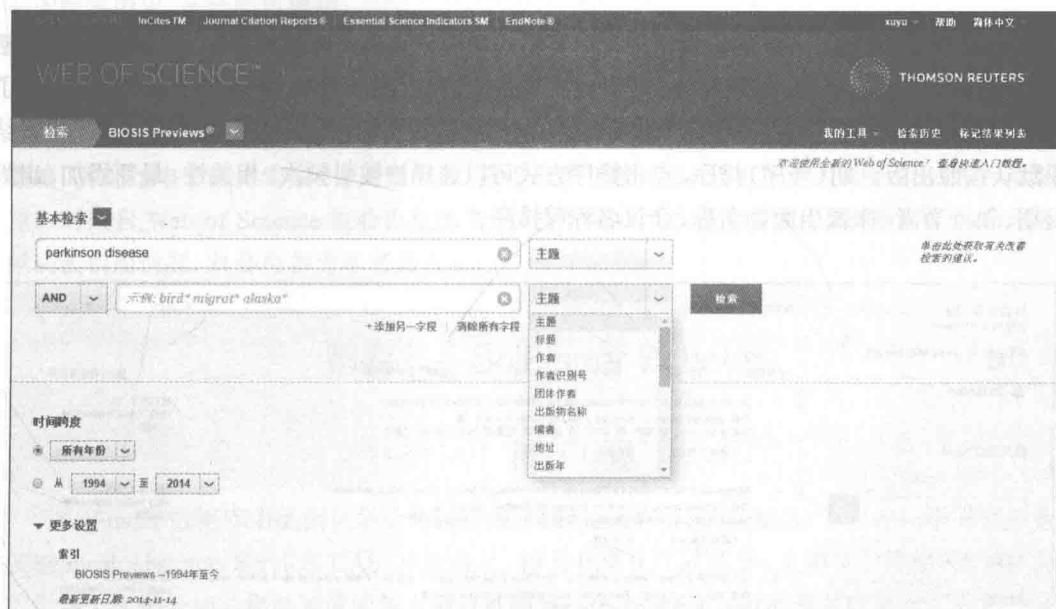


图 5-13 BP 主页面

基本检索时,可直接在检索框中输入检索词或检索式,默认在主题字段检索。主题字段是个复合字段,包括标题、摘要、主要概念、概念代码、分类数据、疾病名称、化学数据、基因名称数据、序列、地理数据、地址年代、设备和方法、器官/系统/细胞器、综合叙词等。也可以在下拉框中选择其他检索字段(见图 5-13)。

点击“添加另一字段”,选择逻辑运算符(AND, OR, NOT),可进行多个字段的复合检索。

例如,要在 BP 中查找中山大学发表的有关鼻咽癌方面的文献,可以在检索框中输入 nasopharyngeal carcinoma, 字段选“主题”,点击“添加另一字段”,逻辑运算符选“AND”,在检索框中输入 Sun Yat-Sen Univ SAME China, 字段选“地址”,点击“检索”按钮,返回检索结果。

2. 高级检索 高级检索时,必须使用两个大写字母表示的字段标识符(field tags),进行布尔逻辑运算,以完成复杂课题的检索。高级检索页面右侧给出了所有可检索字段的标识符。高级检索也支持直接输入检索式的序号进行逻辑组配,还可以对检索年限、语种、出版类型、文献类型、生物分类注释等进行限定(图 5-14)。

例如,要检索埃博拉病毒疫苗的密切相关文献,可直接在高级检索的检索框中输入 TI=(ebola AND vaccin*), 点击“检索”按钮,得到检索结果为 79 篇。

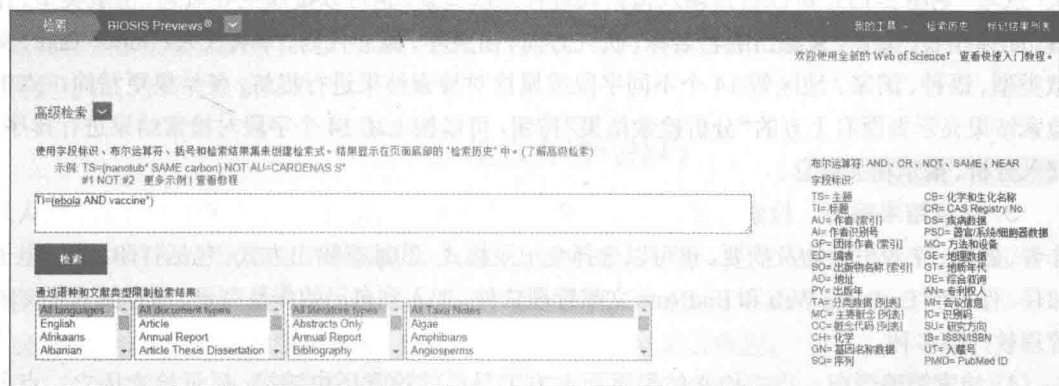


图 5-14 BP 高级检索页面

三、检索结果的处理

1. 检索结果显示 BP 的检索结果有简要格式和全记录格式两种显示方式。简要格式包括篇名、著者、来源出版物、出版时间、被引频次等基本信息(图 5-15), 如果用户所在机构订购了来源出版物的电子版全文, 则显示“出版商处的全文”链接图标, 点击可直接下载全文。检索结果默认按照出版日期(降序)排序, 点击排序方式可以选择按被引频次、相关性、最新添加、出版日期、第一著者、来源出版物名称、会议名称等排序。

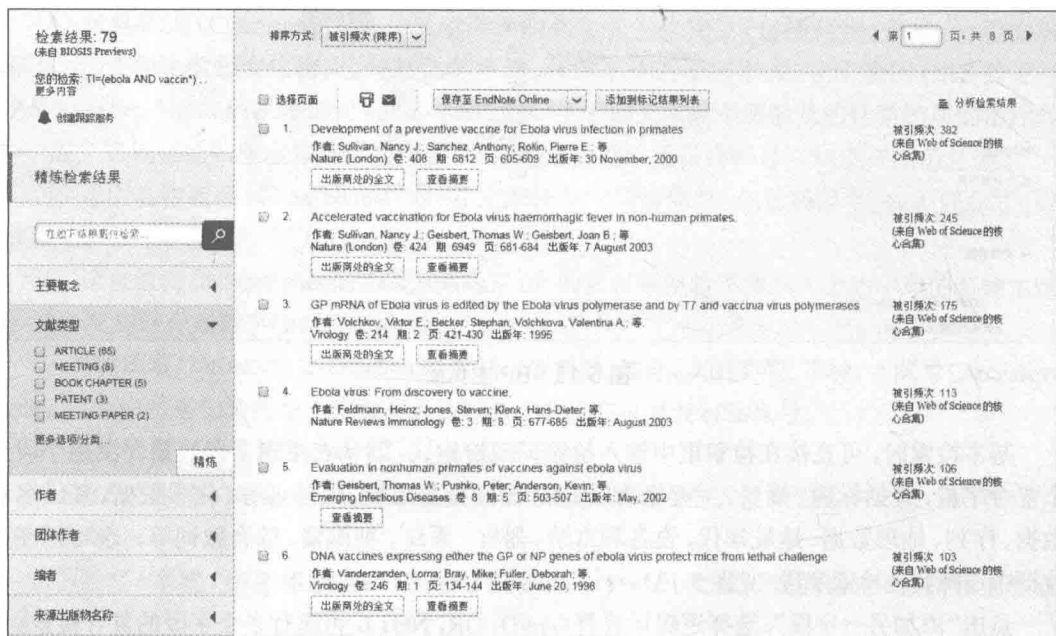


图 5-15 BP 检索结果页面

点击文献篇名进入全记录格式, 包括摘要信息、作者信息、研究方向、主要概念、概念代码、分类数据、疾病数据、化学数据、基因名称数据、文献信息、参考文献篇数、引证关系图、创建引文跟踪、相关记录链接等。页面右侧的“引文网络”详细显示该文献的被引用信息, 包括在 Web of Science 核心合集的被引频次、引用的参考文献、查看 Related Records、查看引证关系图、创建引文跟踪等; “全部被引频次计数”则提供该文献在 WOS 平台所有数据库的被引频次以及在其他引文库(如 BIOSIS Citation Index、中国科学引文数据库等)的被引频次。

2. 检索结果提炼与分析 以简要格式显示检索结果时, 页面左侧提供多种“提炼检索结果”选项(见图 5-15), 可以直接输入检索词进行二次检索, 也可以按照主要概念、出版类型、作者、团体作者、编者、来源出版物名称、研究方向、出版年、概念代码、专利权人、Super Taxa、文献类型、语种、国家/地区等 14 个不同字段或属性对检索结果进行提炼, 使结果更精简。点击检索结果显示页面右上方的“分析检索结果”按钮, 可以按上述 14 个字段对检索结果进行排序, 深入分析、揭示相关信息。

3. 检索结果输出 检索结果输出的步骤: ①选择输出的记录; ②选择输出的字段: 默认为作者、题名、来源出版物及摘要, 也可以选择全记录格式; ③选择输出方式, 包括打印、发送电子邮件、保存到 EndNote Web 和 EndNote 文献管理软件、加入到自己的作品列表、保存到其他文献管理软件等多种。

4. 检索策略保存 点击检索结果页面上方工具栏的检索历史链接, 显示检索历史。点击“保存历史/创建跟踪”按钮, 可以进入检索策略保存页面, 将检索策略保存到 BP 服务器上(需



要先进行个人账户注册), 保存的检索策略可以供下次检索调用, 在检索历史页面点击“打开保存的检索历史”按钮即可调用。

四、个性化功能与服务

BP 可利用 Web of Science 平台提供的个性化服务功能。用户注册后, 可保存感兴趣的检索策略, 并通过电子邮件定期(每周或每月)接收最新文献信息, 还可以使用 RSS feed 跟踪最新报道。对来自 Web of Science 核心合集的文献, 注册用户可以建立引文跟踪, 随时了解感兴趣文献的被引用情况, 具体可参考本书第六章。

第四节 SciFinder Scholar

一、概 述

SciFinder 数据库由美国化学会(American Chemical Society, ACS)下属的化学文摘服务社(Chemical Abstracts Service, CAS)出版发行, 涉及化学化工及药学、生物化学等相关学科, 是目前世界上化学及相关学科领域的最大最权威的数据库, 其前身是世界著名的检索工具刊——美国《化学文摘》(Chemical Abstracts)。SciFinder 还整合了 MEDLINE 数据库的全部文献记录。

SciFinder 由文献数据库 CAPLUS 与 MEDLINE、化学物质数据库 CAS REGISTRY、反应数据库 CASREACT、管制品数据库 CHEMLIST、化学品供应商数据库 CHEMCATS、化学工业札记数据库 CIN、Markush 结构数据库 MARPAT 等 7 个子库组成, 文献信息库、物质信息库及 Markush 结构库的数据每天更新, 其余数据每周更新。

SciFinder 收录的文献数据量极为巨大, 包括 19 世纪以来的全世界 1 万多种自然科学期刊以及 60 多个专利机构发布的 3800 多万条文献记录、8200 余万种有机物质与无机物质数据、6500 余万条物质序列信息、7110 余万个单步或多步化学反应、1350 余万条化学合成研究信息、7500 余万种商业应用化学物质、2600 万种 CAS 独有的登记物质、30 余万种管制化学品目录或物质、1700 余万条化学工业札记、101 余万个可搜索的 Markush 结构与 41 余万个相关专利, 是全球范围内收录期刊文献量最大、专利信息最全的数据库之一。

SciFinder 一般有并发用户限制, 使用前必须先进行用户注册, 注册时需要提供有机构后缀名的邮箱, 用户名与密码的设置也有相应的要求。注册成功后, 即可通过用户名与密码登录 SciFinder 进行检索。

SciFinder 涉及的学科领域非常广泛, 包括化学、药学、生物学和生命科学、医学、化学工程、食品科学、农学、地质、材料、物理等, 提供多学科检索。

SciFinder 包含文献、化学物质、化学反应三种不同类型的文献记录, 提供多种有效的检索途径和强大的检索结果后处理工具, 通过对检索结果的全面分析和深入的数据挖掘, 帮助用户得到可靠、精确的检索结果。

二、检索方法与技巧

(一) 检索规则

1. SciFinder 提供自然语言检索 系统建议在检索时输入能代表研究主题的词或短语, 并在检索词或短语之间插入 of、after、at、with 等连接词, 系统会对短语进行分析, 并通过介词、冠词来判断检索词之间的逻辑关系, 自动给出不同关系的备选答案集。

2. SciFinder 自动进行截词检索 对于用户输入的检索词, SciFinder 自动查询包括该词的单复数、缩写、过去式等各种词性变化形式, 系统不识别“*”、“\$”、“?”等特殊符号。

3. **SciFinder 同义词检索** 采用圆括号“()”表示同义词检索,如输入 treatment of cancer (carcinoma),表示 cancer 和 carcinoma 都检索。系统允许使用 NOT 或者 EXCEPT,表示“除……以外”,例如,要检索“干扰素治疗丙型肝炎,但不包括乙型肝炎”,可输入短语 interferons for the treatment of hepatitis C EXCEPT hepatitis B。

(二) 检索方法

进入 SciFinder 主页,默认为“Explore”检索状态,左栏提供了 SciFinder 的 12 种检索途径,分为文献检索(Reference)、物质检索(Substances)与反应检索(Reactions)三类(图 5-16)。文献检索之下有研究主题、作者姓名、机构名称、文献标识符、期刊、专利和标签等 7 种检索途径;物质检索包括化学结构、Markush 结构、分子式、物性、物质标识符等 5 种检索途径;反应检索提供通过反应结构查找相关化学反应。

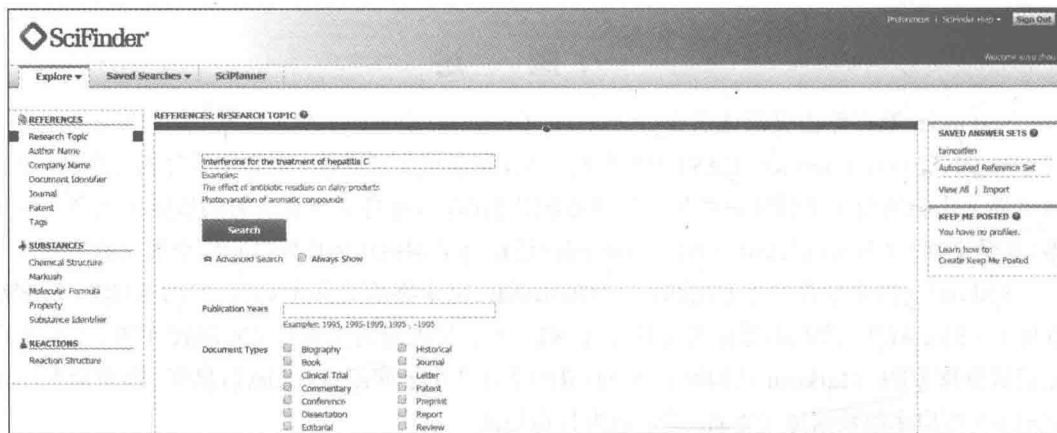


图 5-16 SciFinder 主页面

1. **研究主题检索** 进入 SciFinder,系统默认为研究主题检索(Research Topic)(见图 5-16)。研究主题检索支持自然语言检索,检索灵活、自由,只要在输入框内直接输入代表研究主题的短语即可。为了保证检索效果,建议检索词之间使用介词连接。

例如,要检索“干扰素治疗丙型肝炎”的文献,可直接输入 interferons for the treatment of hepatitis C。

点击检索框下方的“Advanced Search”(高级检索)链接,可提供各种限定条件选项,包括出版年、文献类型、语种、著者、单位等。

2. **作者检索(Author Name)** 尽量输入作者的姓、名、中间名的全称,系统默认勾选“Look for alternate spellings of the last name”选项,即列出姓氏可能的各种拼写形式,以备选择。如果姓名全称不确定,可以仅输入姓或名的首字母。

3. **机构名称检索(Company Name)** 可直接输入机构全称、简称、部分名称或一般称呼,系统会自动识别同一机构的不同表达方式。

4. **文献标识符检索** 可输入特定的文献标识符(Document Identifier)查找相关文献,包括 CAlus 或 MEDLINE 中的文献索取号(Accession Number, AN)、文献号(Document Number)、专利号、PubMed 文献编号(PubMed ID)、数字对象唯一标识符(DOI)。

5. **期刊检索** 可直接输入刊名、称、卷、期、页码、出版年等信息,还可以输入文献题目、作者姓名进行组配检索。

6. **专利检索** 可以在相应的检索框中直接输入 Patent Number(可以是专利号、专利申请号或专利优先申请号)、Assignee Name(专利权人姓名)、Inventor Last Name(专利发明人的姓)、Publication(出版年)等信息。



7. 标签检索 用于重新调用已经保存并做了标签的检索结果。如果在浏览检索结果时,点击页面上方“Tools”(工具)选项下拉菜单中的“Add Tag”链接,将选中的检索结果添加标签并命名。以后只需要点击页面左上方“Explode”下拉菜单中的“Tags”链接,系统会按字顺提供所有已保存的标签,点击选中的标签名即可调用相应的检索结果。

8. 化学结构检索 (Chemical Structure) 可通过化学结构检索相关的化学物质(图 5-17)。检索时先打开化学结构编辑器 (Structure Editor, 图 5-18),可以直接在该编辑器中绘制需要检索的化学结构;也可以导入 (Import CXF) 以 .cxf 格式保存在用户电脑中的化学结构。结构编辑器分为 Java 和非 Java 平台两种,非 Java 平台启动速度更快。化学结构编辑器中绘制的结构可以保存到本地或保存为结构模板。系统提供精确结构检索 (Exact Structure)、亚结构检索 (Substructure)、类似结构检索 (Similarity) 等不同的检索方式选项,点击“Advanced Search”链接,可以对化学结构的特征 (Characteristics)、物质类型 (Classes)、研究类型 (Studies) 等进行限定。

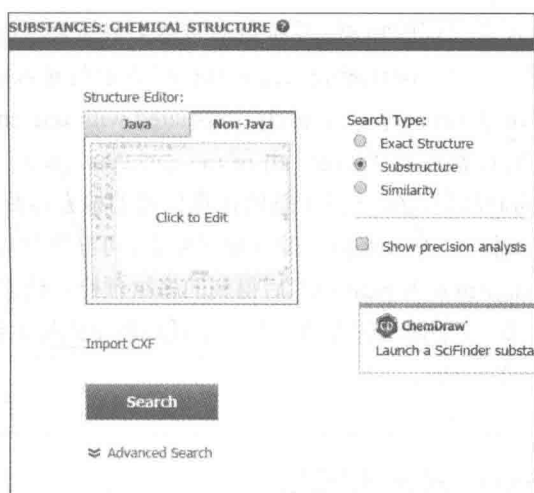


图 5-17 SciFinder 化学结构检索页面

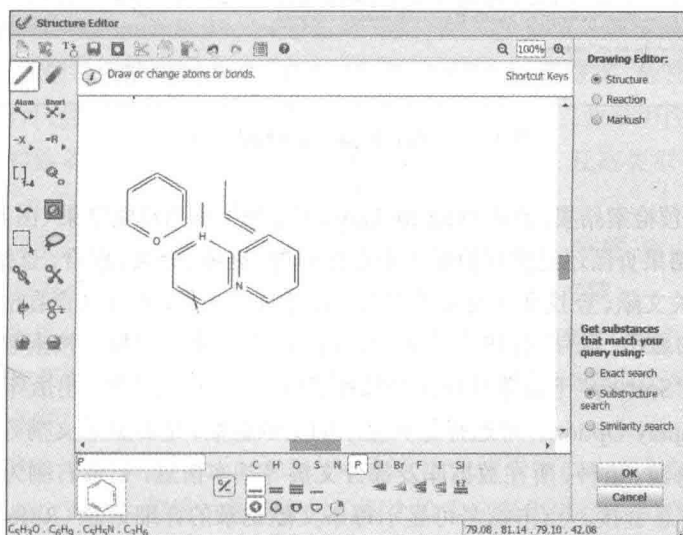


图 5-18 SciFinder 的化学结构编辑器

9. Markush 结构检索 一般用于检索含有 Markush 结构的专利,检索方法与化学结构检索基本相同。

10. 分子式检索(Molecular Formula) 可通过输入化学物质的分子式查找相应的化学物质, 分子式的输入规则与日常书写规则类似, 如果有未知系数的情况, 可以用 x 表示。

11. 物性检索(Property) 可根据实验(Experimental)或预测(Predicted)的物理性质数据检索相应的化学物质。检索时, 先选择是实验数据或预测数据, 再在下拉菜单中选择具体的物理性质选项, 如沸点、密度、电导率、电阻等, 输入对应值或范围, 即可完成检索。

12. 物质标识符检索(Substance Identifier) 可输入 CAS 登记号(CAS Registry Number)、化学物质名称、通用名、商品名、首字母缩写等物质标识符进行检索。

13. 反应结构检索(Reaction Structure) 可通过化学结构检索相匹配的化学反应或化学物质。检索方式与化学结构检索类似, 检索时可打开化学结构编辑器直接绘制化学反应式, 或直接导入以 .cxf 格式保存在用户电脑中的反应式。系统提供两种检索方式, 分别是仅特定位置可变的检索和亚结构检索。

三、检索结果的处理

研究主题检索时, 输入带介词的短语, 系统自动按照多种词间关系进行预检索。预检索提供多组不同的检索结果, 其中“containing...as entered”表示与输入的短语形式完全一致的检索结果, “containing all of the concepts...closely associated with one another”表示所有检索词出现在同一句子或标题中的检索结果, “where all of the concepts...were present anywhere in the reference”表示所有检索词同时出现在一篇文献的任意位置的检索结果, 此外还有不同组合的检索词出现在同一个句子或标题、出现在同一篇文献等结果, 可根据实际需要加以选择。例如, 输入“interferons for the treatment of hepatitis C”后得到的多组预检索结果中, 与输入的短语形式完全一致的检索结果为 52 篇; 3 个词出现在同一个句子或标题中的检索结果为 1738 篇, 这些可视为相关度高的检索结果(图 5-19)。

Select All	Deselect All		References
☒		2 of 12 Research Topic Candidates Selected	
☒		52 references were found containing "interferons for the treatment of hepatitis C" as entered.	52
☒		1738 references were found containing all of the concepts "interferons", "treatment" and "hepatitis C" closely associated with one another.	1738
☒		6806 references were found where all of the concepts "interferons", "treatment" and "hepatitis C" were present anywhere in the reference.	6806
☒		8992 references were found containing the two concepts "interferons" and "treatment" closely associated with one another.	8992
☒		42503 references were found where the two concepts "interferons" and "treatment" were present anywhere in the reference.	42503
☒		4627 references were found containing the two concepts "interferons" and "hepatitis C" closely associated with one another.	4627
☒		9212 references were found where the two concepts "interferons" and "hepatitis C" were present anywhere in the reference.	9212
☒		29191 references were found containing the two concepts "treatment" and "hepatitis C" closely associated with one another.	29191
☒		44085 references were found where the two concepts "treatment" and "hepatitis C" were present anywhere in the reference.	44085
☒		115580 references were found containing the concept "interferons".	115580
☒		9683315 references were found containing the concept "treatment".	9683315
☒		117359 references were found containing the concept "hepatitis C".	117359
Get References			

图 5-19 SciFinder 的预检索结果

勾选感兴趣的预检索结果, 点击“Get References”按钮, 得到检索结果(图 5-20)。除了显示检索结果外, 检索结果页面还提供对检索结果进行分析、提炼、分类、联合, 查找相关物质、查找相关反应、查找相关文献、查找全文及去重等进一步处理的功能, 还可以查看检索历史。

1. 检索结果的显示、保存、打印与输出 SciFinder 的检索结果默认按索取号即入库时间由近及远排列, 通过“Sort by”下拉菜单可选择按作者姓名、被引用次数、出版年和题名排序。点击页面右上方“Display Options”可选择每页显示的文献篇数、是否显示文摘等。每条文献记录提供篇名、著者、来源、语种、所在数据库及部分文摘等基本信息, 记录右侧为该文献的被引用信息链接与物质信息链接。点击篇名可显示每条文献记录的详细信息(Reference Detail), 包括书目信息、主题概念信息、化学物质信息、参考文献信息等。对浏览选中的文献, 点击页面上方的“Get Substance”、“Get Reactions”按钮, 可得到相关的物质信息与反应信息。点击“Get Related Citations”按钮, “Get Citing”链接显示该文的被引信息, “Get Cited”显示该文的参考文献信息。



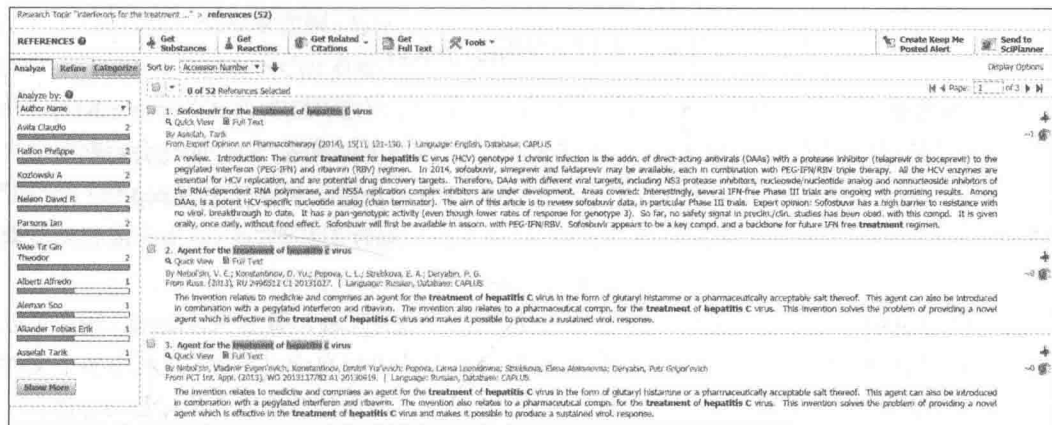


图 5-20 SciFinder 检索结果页面

点击检索结果页面右上方的“Save”按钮,可将选中的检索结果或所有检索结果保存在 SciFinder 服务器中。保存文件后,点击页面左上方主栏目“Saved Searches”下拉菜单的“Saved Answer Sets”按钮,可显示所有已保存的文献、物质与化学反应。

点击检索结果页面右上方的“Print”按钮,选择打印范围、文献格式,并确认是否打印检索历史后,即完成打印。

点击检索结果页面右上方的“Export”按钮,选择输出范围和文件格式,即可将结果输出到本地电脑上。

2. 检索结果的后处理 SciFinder 提供强大的检索结果后处理功能,包括分析、提炼、分类、去重、联合等。其中检索结果的分析、提炼、分类功能位于检索结果页面的左侧区域。

(1) 检索结果分析 (Analyze): SciFinder 可以对三类不同的检索结果集进行分析。文献检索的结果分析项目包括:作者姓名、CAS 登记号、文献类型、单位、CA 概念主题词、期刊、语种、出版年等;化学结构检索的结果分析项目包括:物质角色、元素、工业可用性、有无相关反应文献等;反应检索的结果分析项目包括:催化剂、实验流程、反应步骤、产率、溶剂、作者、机构、文献类型、期刊、语言、出版年等。分析结果可得到柱状示意图。

(2) 提炼检索结果 (Refine): SciFinder 可以对三类不同的检索结果进行提炼,相当于二次检索。文献检索的提炼项目包括研究主题、作者、单位、文献类型、出版年、语种、数据库等;化学结构检索的提炼项目包括化学结构、所含同位素、所含金属元素、工业可用性、物质性质、参考文献等;化学反应检索的提炼项目包括反应结构、产率、反应步骤、反应类型、未参加反应官能团等。

(3) 检索结果分类 (Categorize): SciFinder 提供按照分类标题、分类、索引词等将检索结果从某一大的学科类别具体细化到某一小类,从而快速获得所需要的文献。

(4) 检索结果去重 (Remove Duplicates): 点击检索结果页面上方的“Tools”下拉菜单“Remove Duplicates”链接,可以帮助用户去掉重复的 MEDLINE 记录,也可以点击页面右上角的“Preferences”按钮进行使用偏好设定。

(5) 检索结果联合 (Combine Answer Sets): 点击检索结果页面上方“Tools”下拉菜单的“Combine Answer Sets”链接,可通过将保存过的两个检索结果集进行再次的联合 (Combine)、相交 (Intersect)、去除 (Remove) 等操作,得到新的检索结果集,类似多个检索式的逻辑运算功能。

3. 检索历史 点击检索结果页面左上方主栏目“Saved Searches”菜单下的“History”链接,显示检索历史。如果需要,可以重新调用检索。

四、个性化功能与服务

1. **定题服务** 用户保存检索策略后,系统默认存储在个人账号中,点击“Create Keep Me Posted Alert”按钮可创建定题服务(Keep Me Posted)。定题服务创建后,SciFinder 会定期向用户的注册邮箱中发送更新的检索结果。点击主栏目“Saved Searches”下拉菜单的“Keep Me Posted”选项,可以查看已经设置了定题服务的检索项目,并允许用联合工具进行组配操作或删除检索策略。

2. **SciPlanner** 是一个特定的工作空间,类似模拟合成的区域。通过 SciPlanner,用户可以对多个检索结果进行组织、整合,还可以对反应拟合成路线进行自定义创建,或利用多个文件中的反应、实验步骤、物质和参考文献,设计出满足自己需求的最佳反应或合成途径与方法。

本章小结

本章介绍的四个数据库是生物医学研究领域最常用的文摘型数据库。

PubMed 因其免费开放和收录核心期刊多,更新速度快,回溯年限长,检索方便快捷,是全世界范围内查找医学文献的首选数据库,使用频率最高。PubMed 的自动词语匹配检索和 MeSH 主题词检索是最有特色的检索途径。灵活应用 PubMed 的检索结果过滤、临床查询、引文匹配以及 My NCBI 等工具,能大大提高检索效率。随着 PubMed Central 收录免费电子全文和授权使用机构订购电子期刊数的增加,通过 PubMed 获取文献也愈加方便,而且 PubMed 与 NCBI 的众多事实型、数据型数据库建立了无缝链接。PubMed 的这些优势使其在生物医学文献数据库中始终保持着不可替代的地位。

Embase 是当今世界上最大的生物医学文献数据库之一,检索途径多样、检索功能强大。Embase 整合了 EMBASE 与 MEDLINE 数据库的全部内容,含有 2700 多种 Embase 独有的期刊,尤其是欧洲和亚洲的学术期刊。Embase 的另一个突出特点是它广泛的药学文献的收录。Embase 对药物、疾病以及医疗器械提供了多种专有的副主题词与限定选项,以保证文献的最大可检索性及检索结果的精确性。

BP 作为生物学、实验医学研究的必查数据库,其特点是收录范围广,收录文献类型多,包括专利、会议、专著等。BP 全记录格式提供多个特色字段,如生物分类数据、化学数据、疾病数据等,不仅有助于全面深入地理解文献内容,而且在主题检索时,系统会自动在检索这些特色字段,扩大了检索范围,提高了查全率。此外,整合在 WOS 平台上的 BP,可以与该平台上的 Web of Science 核心合集等数据库实现跨库检索,还可以对检索结果进行深入的分析与提炼。

SciFinder 是世界上最大的化学化工信息数据库,也是生命科学尤其是药学领域重要的文献信息检索工具之一。SciFinder 整合了 MEDLINE 数据库资源,收录最新的高质量文献,相比其他生物医学数据库,SciFinder 拥有更多的期刊和专利文献,不仅提供文献信息,还提供化学合成、化学反应、分子式等事实型数据信息。SciFinder 丰富的检索途径和多样化的分析处理工具能帮助用户得到更全面、更精确的检索结果。

(周旭毓)

思考题

1. PubMed 的检索规则有哪些?什么是自动词语匹配检索?
2. PubMed 有哪几种检索途径?如何应用?

3. 请结合一个检索实例, 写出 PubMed 主题词检索的基本步骤。
4. PubMed 的 My NCBI 有哪些功能?
5. Embase 的高级检索有哪些限定选项?
6. 请比较 PubMed 与 Embase 主题词检索的异同。
7. BP 有哪些特色检索字段? 它们的含义分别是什么?
8. SciFinder 如何进行自动语言检索? 试举例说明。
9. SciFinder 的检索结果后处理功能有哪些?
10. 请结合你的检索体会, 谈谈外文文摘数据库的特点及应用价值。

第六章 外文全文数据库检索

全文数据库是指包括电子图书、电子期刊以及特种电子文献(学位论文、专利、会议文献等)全文的一次文献数据库。

本章主要介绍 Clinicalkey、Journals@ovid Full Text、Wiley Online Library 和 SpringerLink 四种外文全文数据库平台,这些平台中既包括电子期刊、电子图书,也包括临床证据、图片视频等资料,属于多种文献类型全文数据库平台,内容丰富,覆盖面广。此外,还简要介绍 ScienceDirect、CINAHL PLUS with Full Text、ProQuest 以及 Karger 等常用外文全文数据库的使用方法。

第一节 ClinicalKey

一、概述

ClinicalKey(简称 CK)是荷兰爱思唯尔(Elsevier)出版公司于 2012 年 4 月推出的临床医学综合信息服务平台。目前,CK 包含了 12 种文献类型的全文,即 1000 余种医学经典教材或参考书(reference books),其中 95% 被 Doody 核心书目(Core Titles)收录,如《西氏内科学》、《格氏解剖学》和《坎贝尔骨科手术学》等,图书提供 PDF 格式全文;600 余种权威医学期刊,包括 *the Lancet*、*JACC*、*Brain Research* 等顶级期刊;17 000 余个医学或手术视频(medical/surgical videos);300 余个医学操作规范视频(procedures consult videos);220 余万幅医学影像或图片(images),可以在线直接生成幻灯片(PPT)文件;1400 余种疾病主题专论(Topic Pages);850 余个循证医学专论(First Consult monographs),提供疾病全面详尽的信息以支持临床决策;3000 余个药物专论(drug monographs),来源于 Gold Standard's monographs,提供药物详尽信息;4500 余个来源于权威的专业学协会的诊疗指南(practice guidelines);15 000 余份患者教育手册(patient education handouts);12 000 余个来源于美国国立卫生研究院在全球范围的临床试验(ClinicalTrials.gov database);2000 余万条来源于 MEDLINE 数据库的医学文摘。

爱思唯尔合并医学分类法(Elsevier Merged Medical Taxonomy, EMMeT)是爱思唯尔公司自创的医学分类法,它是在美国国家医学图书馆一体化医学语言系统(UMLS)基础上开发,整合了现行多种标准化的医学分类法和主题词表,包括医学主题词表(MeSH)、国际系统医学术语集(systematized nomenclature of medicine-clinical terms, SNOMED CT)、临床药学术标准术语(RxNorm)、国际疾病分类(ICD)以及医学观测指标标识符逻辑命名与编码系统(LOINC)等,EMMeT 拥有 25 万核心临床医学概念,超过 100 万英文医学词汇。ClinicalKey 的内容均根据 EMMeT 进行了深度标引,创建了 CK 的智能内容(smart content)。EMMeT 对检索请求进行更深入的语义和关系分析,以便抓住更加复杂的医学词汇关系。因此,与通常的医学数据库相比,CK 的检索结果相关性更强。CK 智能内容能够理解庞大的医学概念之间的联系,通过把这些关系按照等级进行整理,保证了 CK 能够为用户的检索请求提供具体并且有针对性的答案。

CK 的主要特点包括:内容综合全面,包含了电子书刊、图片视频等 12 种类型的资源,用户在同一平台下无界限获取全面资源;内容权威性较高,所有内容源自爱思唯尔、美国国立卫生研究院和专业学协会;用户界面简单直观,方便友好,个性化检索功能、定制化功能、PPT 制作

工具和邮件分享功能进一步节省检索时间,提高信息分享效率;检索快速准确,所有内容依据 EMMeT 进行标引,支持语义检索,显著减少用户在海量检索结果中反复筛选的时间,更快捷提供最准确相关的答案。

2014 年 9 月,CK 平台升级到 2.0 版,检索界面和功能得到进一步优化和提升。内容上新增 1400 多个医学主题,对 1400 多种疾病进行了全面概述,包括风险因素、临床表现、治疗等可支持快速床旁决策的内容,以及专科医生进行深入研究所需要的资源链接。患者教育手册定制允许用户根据患者情况选择打印语言和字体大小,也可以添加医嘱和医生联系信息。移动设备优化意味着用户可以在任何设备上(包括 PAD 或手机)访问 ClinicalKey,随时随地查找所需答案。幻灯片制作简化了程序,用户日后分享、操作和存储内容更加方便,用户可以创建、命名和管理多个演示文档。

ClinicalKey 信息服务平台的网址为: <http://www.clinicalkey.com>, 其主页如图 6-1 所示。

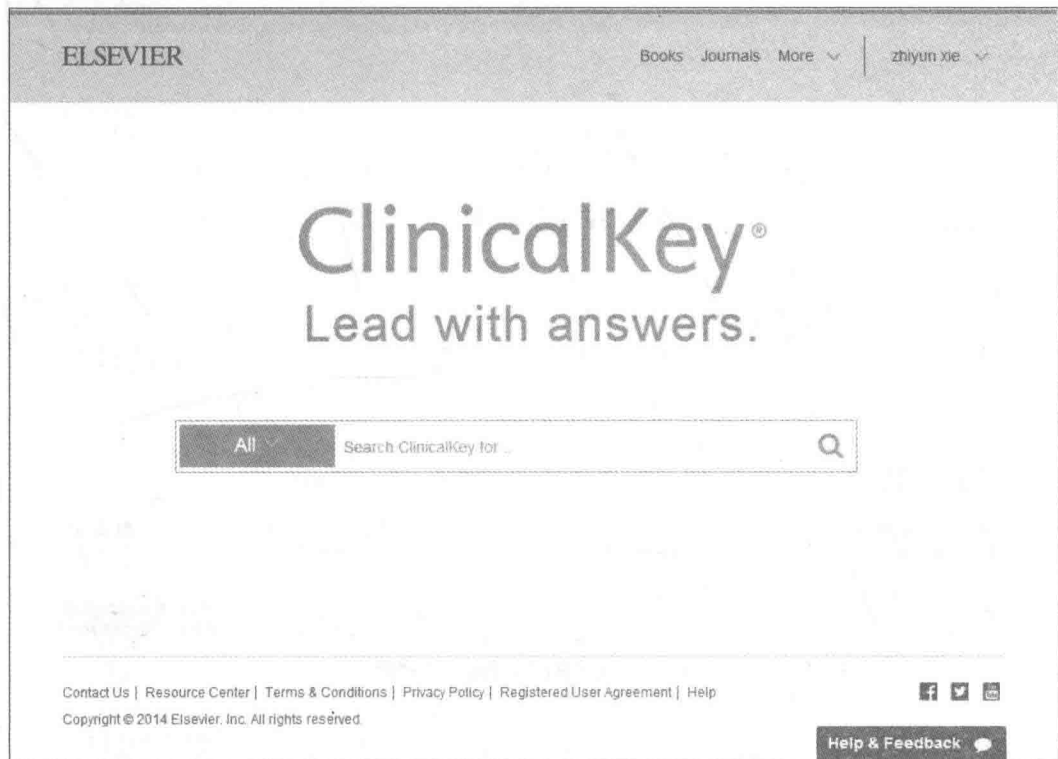


图 6-1 ClinicalKey 主页

二、检索方法与技巧

ClinicalKey 的使用方法主要有检索和浏览两种方式。ClinicalKey 的检索界面简单明了,使用方便,并不涉及复杂的检索规则和语法,但由于内容均经过医学分类法处理和标引,因而检索结果比较准确。因此,这个工具特别适合工作繁忙的临床工作者在工作中快速准确地查询所需要的临床信息,解决医疗实践中的问题,而使用之前也无需太多的培训。

(一) 检索

进入 ClinicalKey 主页(见图 6-1)后,即可见类似谷歌的检索输入框,可输入相应检索词进行检索。输入框左侧的下拉菜单可选择特定的文献类型进行限定检索,系统默认为 All,即全部文献类型。主页右上方有浏览功能菜单,可按文献类型浏览 CK 内容。主页右上方还有个人账号注册和登录选项。个人账号登录后可以进行个性化操作,包括查看已保存检索结果、检索历

史、幻灯片制作、个人账号设置以及退出账号等。在任何界面点击 ClinicalKey 图标即可返回主页。

在检索输入框中,可输入完整或部分单词或短语、缩略语(如输入 CHF 将检索 congestive heart failure 的文献)、作者姓名、图书或期刊名称以及美国国家医学图书馆标准题录格式(如可输入“Am J Cardiol. 2011 Dec 1; 108(11): 1614-9”)进行检索。

在输入检索词的过程中,CK 还提供“自动建议”功能,主动为用户推荐检索词或提供检索词建议。当用户输入检索词的前几个字母的时候,“自动建议”功能即开始工作,按照用户输入的字母组合,在检索框的下方弹出建议检索词列表,列表包括建议词、相关检索词、图书或期刊名称以及可能的作者姓名等四类。随着用户输入的检索词的进一步完善,“自动建议”功能也同步调整或提炼建议词列表(图 6-2)。

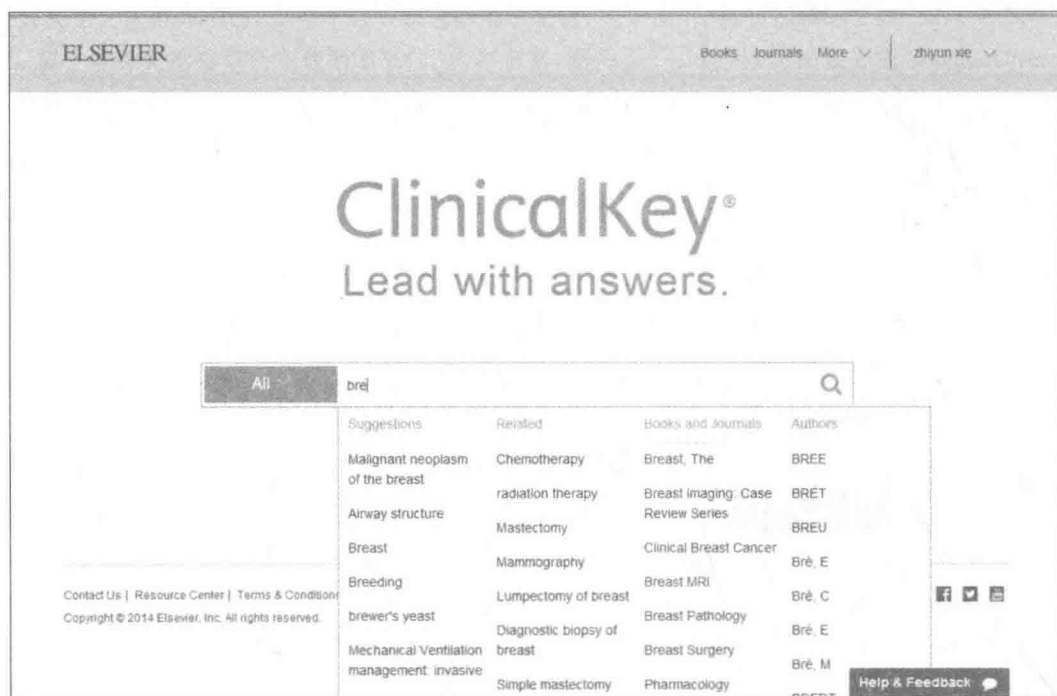


图 6-2 CK 的“自动建议”功能

在建议检索词列表中选择最合适的检索词,点击选中的检索词即可完成检索过程。当然,用户也可不理睬建议词列表,直接输入检索词,点击检索按钮完成检索。检索结果按照相关度由高到低排列。

(二) 浏览

除检索功能之外, ClinicalKey 还提供按文献类型浏览其内容。浏览功能菜单在所有界面的顶端,提供按 Books(图书)、Journals(期刊)、Drug Monographs(药物专论)、Guidelines(诊疗指南)、Patient Education(病患教育手册)、Multimedia(多媒体)等六大类分别浏览。

点击 Books 链接即可进入图书浏览界面(图 6-3),系统按书名的字母顺序罗列了 CK 全部图书的书名。界面右侧是系统推荐的新书封面和名称。

为了快速获得需要的图书,可在 Browse Books 标题下的检索框中输入完整或部分书名进行检索,快速定位需要的图书;或者通过医学专科过滤器(Filter)选择感兴趣的专科,获得相应专科的图书,医学专科可以复选多项。

点击需要的书名链接,即可进入图书目次页;点击章节名称链接即可打开图书阅读正文(图 6-4)。



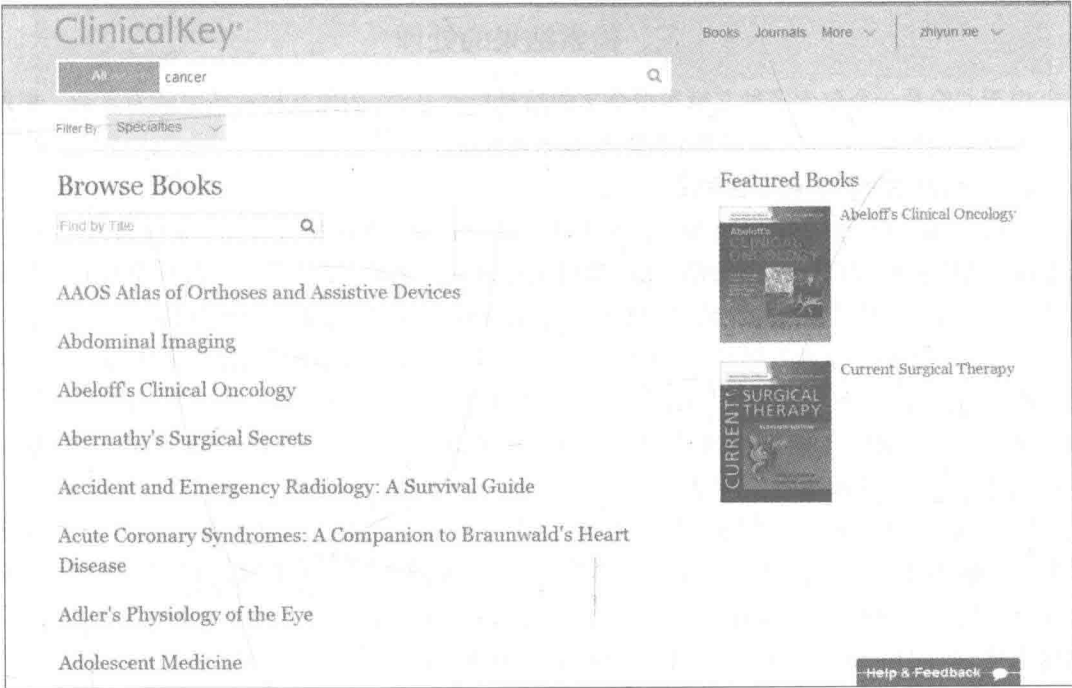


图 6-3 图书浏览界面

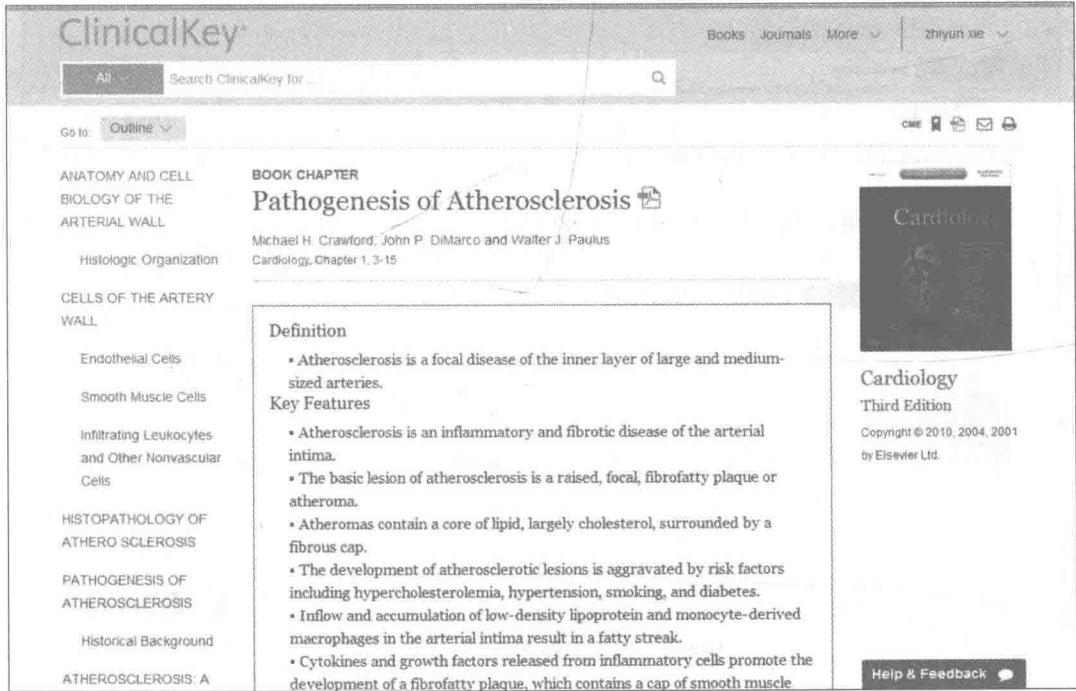


图 6-4 图书阅读界面

点击 Journals 链接即可进入期刊浏览界面，系统按刊名的字母顺序罗列了 CK 全部期刊刊名。界面右侧是系统推荐的新刊封面和名称。同图书浏览一样，也提供刊名或刊名词检索特定期刊，或医学专科过滤器筛选期刊。选定期刊后，可通过年、卷、期浏览阅读期刊文章。

点击 Multimedia 链接可浏览医学图片和视频。系统提供医学专科过滤器和媒体类型（图片或视频）过滤器筛选阅读医学图片或视频。

其他文献类型的浏览方法与上述类型类似，只是过滤器有些差别。

三、检索结果的处理

由于 CK 是一个包含多种文献类型全文的信息检索平台,因此其检索结果较为复杂。根据输入的检索词性质不同,可以得到不同的检索结果和展示。

(一) 检索结果为疾病主题专论

目前, ClinicalKey 提供 1400 余个疾病主题专论,对 1400 多种疾病进行了全面概述,包括流行病学、风险因素、临床表现、诊断、治疗等可支持快速床旁决策的内容,以及特定专业问题解决方案和疾病相关药物的链接,还包括专科医生进行深入研究所需要的资源链接。

当在主页检索框中输入的是 1400 多种疾病之一时,系统的检索结果将首先默认显示该疾病的主题专论页面,如图 6-5 所示。页面左侧列出了有关该疾病的全面概述,这种全面概述可来自多种文献源。本例中乳腺癌的全面概述给出了三个来源: *Ferri's Clinical Advisor 2015* (《费里临床顾问 2015》)、*Goldman's Cecil Medicine* (《希氏内科学》) 和 *First Consult* (《循证医学专论》)。通过阅读这些内容,可以协助临床医学快速回顾和了解疾病的基本情况和最新进展,为临床诊疗决策提供参考和指导。很多细节内容可以通过超链接进一步深入了解。页面右侧给出了有关该疾病的专业观察 (Specialty Views)、相关病患教育手册 (Related Patient Education)、最新文献报道 (Latest Articles) 和相关诊疗指南 (Related Guidelines), 点击相应链接即可获得详细资料。

当用户不希望阅读疾病主题专论部分,则可点击页面上端的 Show all results 链接进入常规检索结果页面。

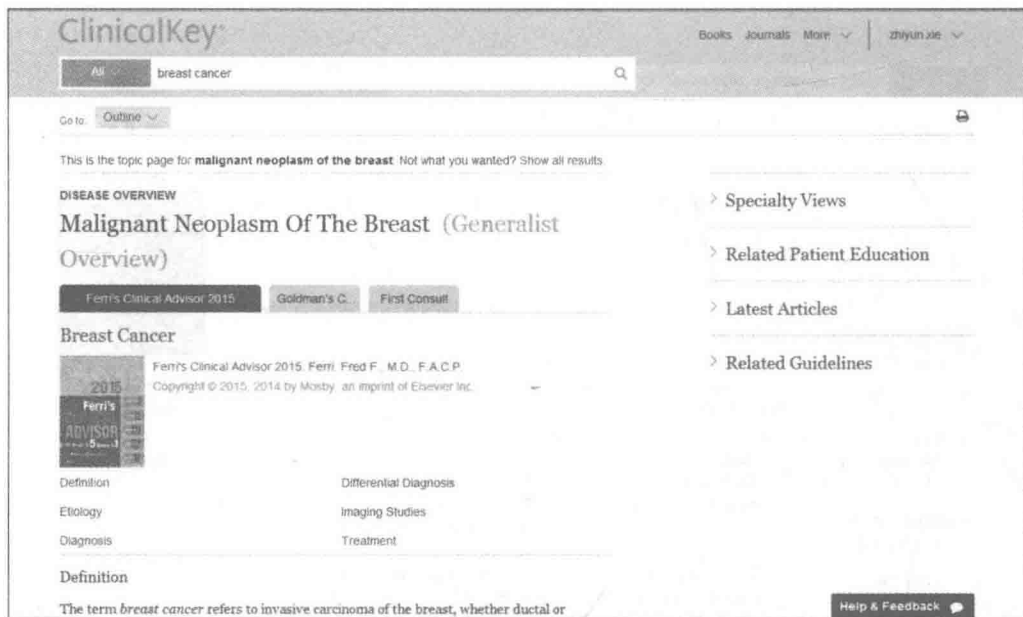


图 6-5 疾病主题专论 (topic page)

(二) 常规检索结果页面

当用户输入的检索词并非 1400 个疾病主题时,系统将进入常规检索结果页面 (图 6-6)。

页面上端首先报告文献检索总量,这个总量包含各种文献类型。检索总量左边的复选框点选后,可对各种类型的检索结果进行人工挑选,以备统一输出。检索结果默认按照相关度排序,也可点击 Sort by (排序) 菜单选择按出版日期排序。

在检索结果总量的上方有 4 个过滤器可对检索结果进行限定,进一步提炼检索结果。这 4 个过滤器分别是: 资源类型 (Source Type) 过滤器 (可选择特定文献类型对检索结果进行提炼)、研究类型 (Study Type) 过滤器 (可选择系统性综述 (Systematic Reviews)、荟萃分析 (Meta



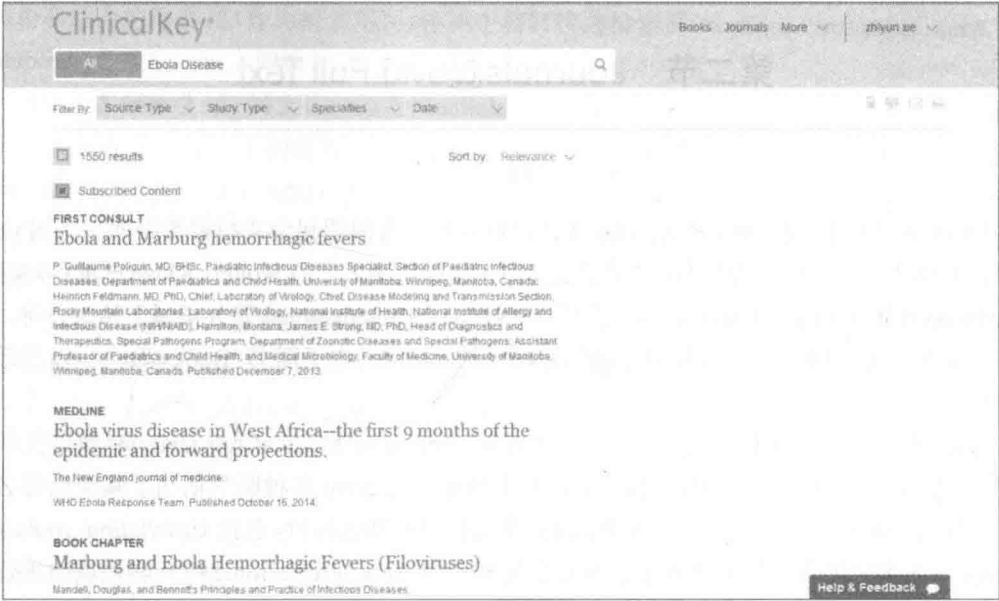


图 6-6 常规检索结果页面

Analysis)、随机对照试验(Randomized Control Trials)和叙述性综述(Narrative Review)、医学专科(Specialties)过滤器(可选择特定医学专科提炼检索结果)和出版日期(Date)过滤器(可选择最近半年、1年、18个月、2年或5年的文献)。

在浏览阅读某本书、某本刊或某篇文章时,可在页面上端的检索框中输入检索词,并限定在某本书、某本刊、某一期甚至某篇文章内检索。

对于选中的检索结果,系统提供保存(save)、打印(print)、电子邮件(Email)和添加到幻灯片制作工具(Add to Presentation,仅针对图片检索结果)四种输出方式。

四、个性化功能与服务

ClinicalKey 允许并鼓励用户在使用这个信息平台之前申请注册个人账号并登录后再开始使用这个系统,这样才能获得 CK 为注册账号提供的个性化功能与服务。

(一) 在线查看保存的检索结果

只有当用户注册账号并登录后,才能将检索结果进行在线保存。结果保存在 CK 为每个账号在服务器上建立的独立空间中。文献保存过程中,用户可以通过“标签”功能对保存文献进行分类、管理和组织,以备日后快速访问和使用。

(二) 检索历史

用户账号登录并开始检索后,CK 即开始跟踪用户的检索行为,记录用户的检索过程,自动形成用户的检索历史。在检索历史页面,用户可以点击任何一个之前的检索词(检索式),系统将重新检索,并获得新的检索结果。用户还可以将最常用的检索历史保存起来,以方便日后使用。在后续的使用过程中,系统会自动刷新检索结果,提供最新的、最相关的文献。

(三) 幻灯片制作

系统允许用户将图片类型的检索结果输出添加到系统内嵌的幻灯片制作工具中,以方便用户随时制作幻灯片,与其他同行或用户进行学术交流。CK 允许用户按照自己的意愿建立、命名和管理幻灯片。最后通过点击 Export 按钮生成可以播放的 PPT 文件。

(四) 定制病患教育手册

允许用户根据患者情况选择打印语言和字体大小。也可以添加医嘱和医生联系信息。

(谢志耘)

第二节 Journals@ovid Full Text

一、概 述

Journals@ovid Full Text 是美国 Ovid 公司向全球推出的期刊全文检索数据库。美国 Ovid 公司是全球著名医学信息服务商, 隶属于全球四大出版集团之一的 Wolters Kluwer(威科)集团, 与 Lippincott Williams & Wilkins(以下简称 LWV)、Adis International 等公司属于姊妹公司。通过其检索平台 OvidSP, 可以检索电子期刊(Journals@ovid)、电子图书(Books@ovid)以及近百种数据库。

Journals@ovid Full Text 提供访问包括 LWV、American Medical Association(美国医学协会)和 Adis International 出版的医学药学核心期刊在内的 3000 多种医学期刊。其中最著名的期刊专辑是 LWV 出版的生物医学期刊, 内容涉及护理和临床各科, 包括 *Circulation*、*Annals of Surgery* 等著名的期刊。另外药理学和治疗学期刊专辑 Adis International 中药学领域的重要期刊也通过 OvidSP 向用户提供使用。

除电子期刊外, OvidSP 也提供了电子书(Books@Ovid, 提供包括 LWV、Adis International 及其他出版社出版的 1000 多本电子书), 以及生命科学、医学、农业、生物、工程技术等领域上百种数据库。著名的数据库包括 MEDLINE、BIOSISPreviews、Embase、PsycINFO、EBMReviews(循证医学数据库)、CABAbstracts、Agris、Agricola、Adis R&D Insight、三维视频解剖学数据库 Primal 3D 等都通过 OvidSP 提供检索。

二、检索方法与技巧

OvidSP 的所有数据库使用同一检索界面提供检索。检索平台界面清晰明快, 支持智能检索功能。主要的六种检索途径分别为基本检索(Basic Search)、题录检索(Find Citation)、规范化主题词检索(Search Tools)、字段检索(Search Fields)、高级检索(Advanced Ovid Search)、组合检索(Multi-Field Search)。检索主界面由检索输入区、输出方式设置区、结果显示区三部分组成。检索输入区界面如图 6-7 所示。

图 6-7 OvidSP 检索途径、检索条件输入区页面

(一) 基本检索

OvidSP 基本检索支持“自然语言检索功能”(natural language processing, NLP)。直接在基本检索框里面输入相关关键字或一句话即可, 系统将自动提取检索用词。例如在基本检索框中输入 warfarin for treatment of dvt, OvidSP 过滤器将自动提取其中的主要词语(包括名词、名



词短语和适当的形容词)作为检索用词,本例中将转换为检索词:warfarin、treatment、deep vein thrombosis,各检索用词之间执行 AND 运算。

系统可扩展派生词、同义词。勾选“Include Related Terms”复选框,将自动扩展该检索词的相关派生词、同义词、不同拼写法等。如上例中检索词语扩展为:warfarin, warfarinsodium, coumadin; treatment, therapeutic aspects, disease management, therapy; dvt, deep vein thromboses, deep vein thrombosis, deep venous thromboses。每组扩展词之间执行 OR 运算。通过扩展相似词语,提高文献查全率。

系统提供了查找多媒体文献的功能。勾选“Include Multimedia”,系统可检索出多媒体文献。在检索限制(Limits)区域,点击“Limits”链接后可以直接设定检索年限、文献类型、是否有全文等,点击“Additional Limits”,增加更多限定条件。

(二) 题录检索

通过“Find Citation”,输入已知的题录信息,如篇名、刊名、作者姓名、期刊的年、卷、期、起始页等,可快速查找特定文章。需要注意的是刊名必须输入全称,可输入刊名起首部分,选择“Truncation”截词检索。著者姓名的输入格式为“姓”全称在前,“名”首字母在后,也可选择“Truncation”截词检索。

(三) 字段检索

输入检索词或词组,选择单个或多个字段,点击“Search”显示检索结果。OvidSP 提供了常用的检索字段 Author、Abstract、Keywords、Full Text 以及全字段“All”,并可以添加进“My Fields”做个性化设置。点击“Display Indexes”可显示该检索词的相关索引词,帮助正确选词。

(四) 高级检索

OvidSP 平台为专家级用户提供支持命令行的精确检索技术,提高查全率和查准率。

OvidSP 在高级检索界面下规定了多种检索运算符和检索规则,编制检索策略时必须正确运用这些运算符和运算规则:

1. 邻近算符“adjn”表示两个检索词之间最多允许插入 $n-1$ 个单词,且词序可互换,例如 physician adj3 relationship。

2. 截词符 包括无限截词符“\$”、有限截词符“\$n”。无限截词符“\$”用于单词词尾,替代任意多个字符,例如 radiolog\$。有限截词符“\$n”,用于单词词尾,可替代 n 个字符,例如 dog\$1。

3. 通配符 包括强制通配符“#”和可选通配符“?”。强制通配符“#”可用于单词的词中或词尾,替代一个字符,并且此位置必须出现字符,例如 wom#n。“?”为可选通配符,可用于单词的词中或词尾,替代一个字符,此位置可以出现或不出现字符,例如 colo?r。

4. 词频限制符号 “/freq=n”,系统允许限定检索词在指定字段中出现的次数(即频率),限定方法为“检索词.字段名缩写./freq=n”。例如 blood.ti./freq=3,可检出篇名中 blood 至少出现 3 次的文献。

5. 特殊字符检索 对于特殊字符串,用半角引号括起来进行检索。例如 "3".vo 检索卷号为 3 的文献。

OvidSP 平台高级检索包括以下检索入口:

1. 关键词检索(Keyword) 可以输入检索词、词组或含逻辑运算符的检索式,系统直接执行检索。默认检索字段包括题名(Title)、文摘(Abstract)、全文(Full Text)和图片说明(Caption Text)四个字段。输入多个单词以空格分隔,系统默认为词组检索。但如果检索短语中含有禁用词(如 and, or, not, .ti. 等),必须用半角的双引号括起短语,否则系统将把禁用词作为逻辑运算符处理。例如 "acute and chronic low back pain"。字段限定符为“. 字段名缩写”。例如 adiponectin .ti, 检索题名中出现 adiponectin 的文献; adiponectin.ti, nm, 检索题名或物质名字段出现 adiponectin 的文献。

2. 著者检索 (Author) 输入完整的著者姓名, 姓全称在前, 名首字母在后。例如 milleram, 也可只输入著者的姓, 则系统检索出所有该姓氏的著者发表的文献。

3. 题名检索 (Title) 指在文献题名中检索, 可以输入单个检索词、词组或检索表达式, 表示检索“Title”字段中含有检索词或词组的文献。

4. 刊名检索 (Journal) 输入完整的刊名全称, 检出发表在该刊上的文献。也可输入刊名全称起始部分的单词, 点击“Search”, 则系统检索出所有以输入的单词开头的期刊刊载的文献 (注意不能输入刊名缩写)。

(五) 组合检索

系统提供多个字段的组合检索, 以提高检索效率。各字段检索结果之间进行 AND、OR、NOT 的逻辑组配。

(六) 主题词检索

为了便于用户使用主题词进行检索, OvidSP 平台设置了规范化主题词检索工具, 帮助用户了解规范化主题词的定义、注释、历史变更、适用范围, 确定规范化主题词、副主题词等信息。通过主题词配对功能“map term”, 将自由词转换成为相关的主题词, 并以相关性排序。主题词的树状结构“Tree”展示该词上下级层次结构, 便于选择合适的主题词。选择主题扩展检索 (Explode) 后, 检索结果包含该检索词及其下位词 (narrow terms) 的全部文献, 提高文献查全率。选择相应副主题词 (Subheadings) 用于检索该主题词某一方面的文献。

需要注意的是, OvidSP 平台中不同的数据库系统将选择其对应的不同的控词词表。例如 MEDLINE 数据库使用 MeSH 词表, Embase 数据库使用 Embase 的控词词表, 而 Journals@Ovid 或 Books@Ovid 数据库不支持控词词表, 也不显示规范化主题词检索工具 Search Tools 选项。

三、检索结果的处理

每次检索完成, 若检索结果不为零, 则以题录方式列出。在检索结果的左侧, 系统提供进行二次检索、检索结果根据年限、期刊、作者等的筛选功能; 每篇文献题的右侧则列出检索系统提供的该文献的相关信息, 如全文、摘要、参考文献等的链接。

为避免检索出大量不相关的文献, 系统按照检索词命中的次数 (Count)、检索词出现的频率 (Frequency)、重要性等相关算法, 按照关联度的大小对文献星级评定, 其中 5 星级文献相关度最高 (图 6-8)。

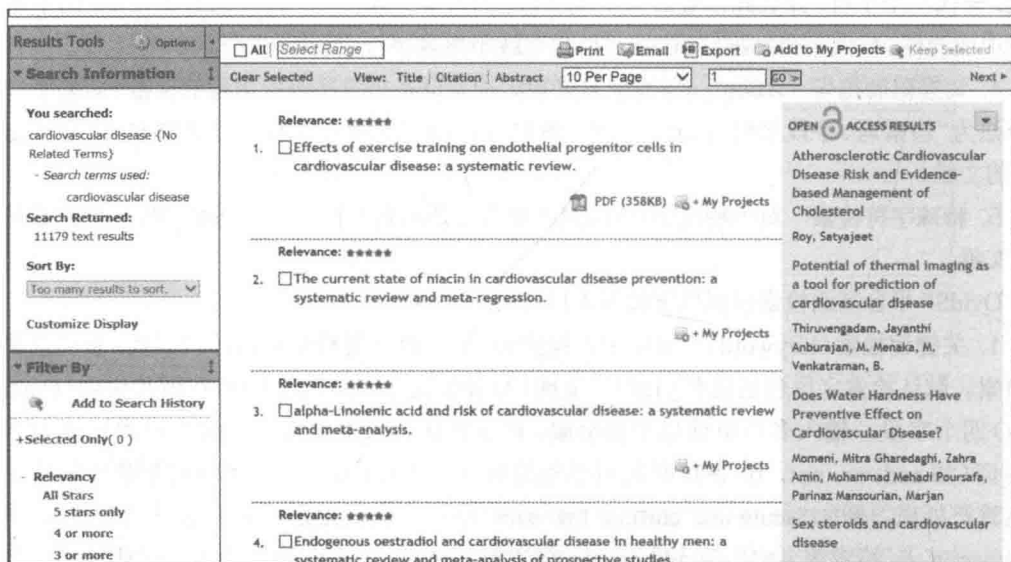


图 6-8 OvidSP 检索结果显示页面

通过“Search History”可以浏览检索历史,还可以进行检索式的组配,即勾选其中的几个检索式,用“**And**”、“**Or**”进行组配形成最终的检索式。

完成检索后,浏览检索结果,可标记重要文献,并对标记的检索结果进行保存、打印或发送电子邮件。文献全文应将全文链接打开后逐篇单独输出。

四、个性化功能与服务

(一) 保存检索策略

通过个人账户(My Account),可以对检索策略进行保存。第一次使用个人账户,需要进行简单的注册,再次使用只需输入设定的用户名和密码,便可显示所有保存过的检索策略,可进行重新检索或者查看更新文献。选择自动定题提示功能“**AutoAlert**”,系统将根据用户设定的时间自动对已存档的检索策略重新检索,把最新的文献发到指定的电子邮箱,以跟踪了解课题最新进展(图 6-9)。

图 6-9 OvidSP 自动提醒功能设置界面

利用 E-mail Jumpstart 为任何已被存档的检索策略建立超链接,只需要点击电子邮件超链接,直接显示更新后的检索结果。

(二) RSS 定制功能

Journals@Ovid 平台支持多种 RSS feed 工具,通过任何 RSS 聚合软件(如 Sharp Reader, NewzCrawler, Feed Demon)定制后,可快速获取期刊电子目录 eTOC 等更新信息。具体使用方法为:选择数据库并进行检索,点击进入 RSS 链接,复制 RSS 种子链接快捷方式,粘贴至 RSS 客户端软件如 Sharp Reader 等。只需点击 RSS 客户端软件的“Refresh”,即可收取到数据库中最新检索结果的链接。

(三) eTOC(electronic table of contents, 电子目录)提醒服务

选择需要订制目录的期刊,设置订阅信息的格式和 E-mail 地址,当期刊目录内容更新时系统自动将更新后的目录信息发送至电子邮箱。

(四) 灵活多样的在线帮助

包括专家问答(Ask a Librarian),启动专家提问系统,填写基本联系方式、描述检索中所遇

到问题,系统将问题描述发送至图书馆参考咨询馆员,由参考咨询馆员解答后予以回复。

(孔瑞珍)

第三节 Wiley Online Library

一、概 述

Wiley Online Library 是 Wiley-Blackwell 在网上推出的电子文献信息检索平台。美国 John Wiley 公司是一家有着 200 多年历史的国际知名的专业出版机构,创建于 1806 年,在生命科学、医学、化学以及工程技术等学科领域的文献出版方面颇具权威性。2007 年,John Wiley 与另一家国际知名的学术出版机构 Blackwell 公司合并,称为 Wiley-Blackwell。

Wiley-Blackwell 的电子资源通过 Wiley Online Library 平台提供浏览访问及检索服务。2014 年,Wiley Online Library 平台上可访问的学术期刊有 2314 种,涉及生命科学与医学、数学统计学、物理、化学、地球科学、计算机科学、工程学、商业管理金融学、教育学、法律、心理学等多个学科领域。此外,Wiley Online Library 平台上还包括 16 177 种图书(包括 Books 和 Book Series)、18 种实验室指南(Lab Protocol)、16 种数据库(Databases)。订购了上述资源的机构用户都可以实现全文下载与检索。本节主要介绍 Wiley Online Library 平台上电子期刊全文数据库的使用方法。

二、检索方法与技巧

进入 Wiley Online Library 主页(图 6-10),系统提供浏览和检索两种检索功能。

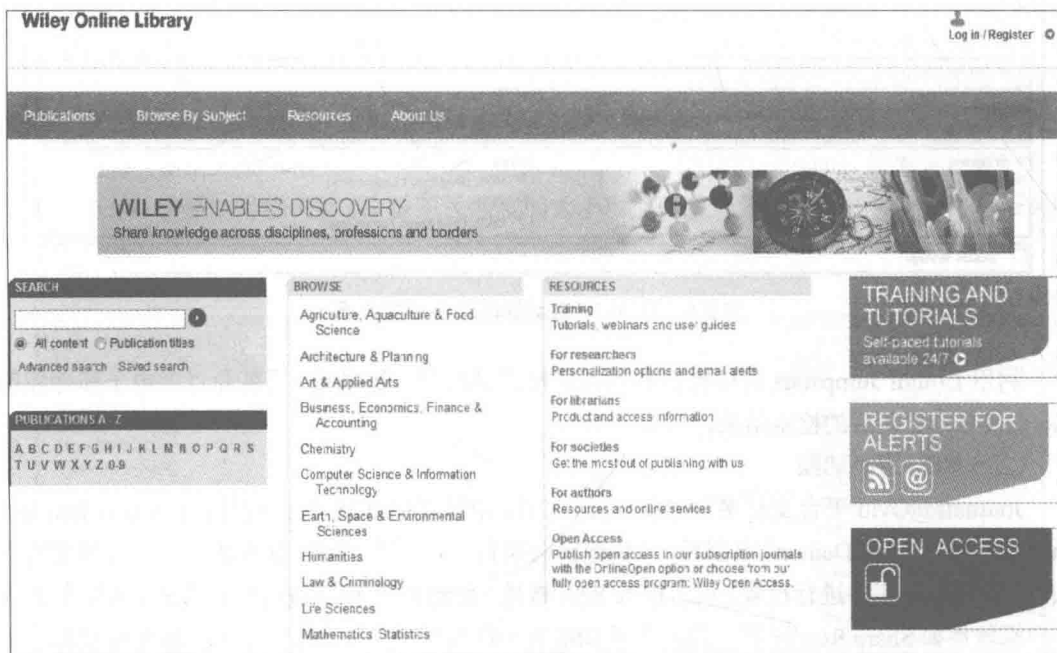


图 6-10 Wiley Online Library 平台主页

(一) 浏览

1. 按主题浏览书刊 在 Wiley Online Library 主页面,根据学科主题浏览(Browse By Subject)栏目之下列出了 17 个学科主题,点击进入感兴趣的学科主题页面,每个学科主题之下又细分为多个次级学科主题类目,继续点击感兴趣的类目,进入该次级学科主题页面,其下按字顺列出



了该次级学科主题的资源一览表,包括期刊、在线图书、参考工具书等。选择感兴趣的期刊,点击即进入每种期刊的内容浏览页面(以 *British Journal of Dermatology* 为例,图 6-11)。页面上方有期刊工具、期刊菜单、特殊功能选项。期刊工具包括设置电子邮件提醒(Get New Content Alerts,即通过邮件发送期刊的最新目次)、RSS 订购、将该期刊保存到我的档案(Save to My Profile,可建立个人感兴趣的期刊目录)、获取该期刊的样本(Get Sample Copy)、推荐给您的图书管理员(Recommend to Your Librarian)等。期刊菜单包括卷期导航(FIND ISSUES)、文献导航(FIND ARTICLES)、获取(GET ACCESS)、贡献者服务(FOR CONTRIBUTORS)、期刊介绍(ABOUT THIS JOURNAL)等。卷期导航(FIND ISSUES)下为“Current Issue”显示最新一期的文献;“All Issues”显示全部文献。文献导航(FIND ARTICLES)下为“Early View”显示已经在 Wiley Online Library 平台在线发布但尚未印刷出版的最新文献。获取(GET ACCESS)显示订购或更新的文献。

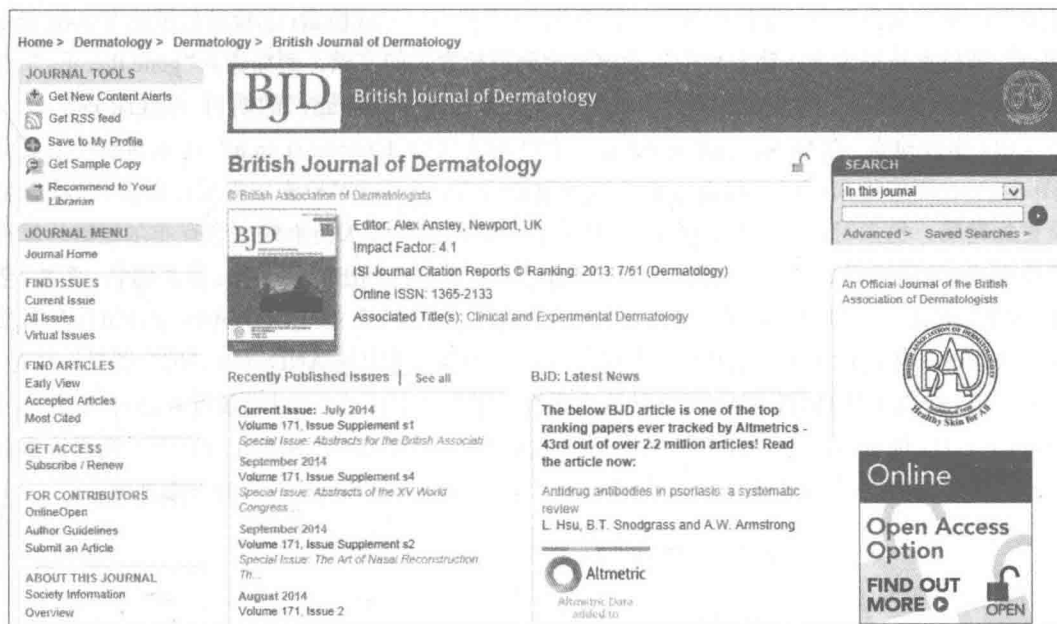


图 6-11 Wiley Online Library 平台的期刊浏览页面

页面中部为期刊封面、刊名、期刊影响因子及卷期列表。点击相应的卷期即可显示本期的文献。

期刊浏览页面的右侧为检索栏,包括:①一般在所有期刊中检索(All content):在 Wiley Online Library 平台检索;②期刊检索(Publication titles):查找某种期刊;③在这个期刊中检索(In this journal):输入检索词,点击^①,即在该期刊中进行二次检索,例如在浏览 *Journal of Clinical Ultrasound* 时,想查找该刊中有关“胎儿心脏超声”的文献,可以在检索框中输入 heart AND (fetal OR fetus),点击“检索”;④引文检索(By Citation):输入已知的卷、期、页码来查找某一特定文献。

2. 按字顺浏览书刊 在平台主页左侧可直接按字母顺序浏览出版物,也可以通过点击“Publications”,打开全部资源列表界面,在此界面右侧有按文献类型的筛选的链接及各类型文献的数量,直接点击文献即可打开相对文献类型的目录。

(二) 检索

Wiley Online Library 平台检索功能强大,支持布尔逻辑检索、截词检索、短语精确检索、邻近检索等多种检索技术,有基本检索、高级检索两种检索途径。

1. Wiley Online Library 平台的检索规则

(1) 支持 AND, OR, NOT 布尔逻辑检索。直接输入几个词, 系统默认词之间为 AND 逻辑关系; 可以用“,”代替 OR 逻辑运算。

(2) 支持截词检索, 截词符为“*”。

(3) 支持短语精确检索。在短语词组上加双引号“”, 系统会将短词视为一个不可拆分单元进行检索。

(4) 支持邻近检索, 邻近算符为 NEAR 及 NEXT。“NEAR/n”表示两个检索词之间最多可以插入 n 个单词。“NEXT”表示两个检索词必须相邻, 但前后顺序不限。

2. Wiley Online Library 平台的检索途径

(1) 基本检索: Wiley Online Library 平台主页 (见图 6-10) 左上方有基本检索入口: ①检索范围默认为所有内容, 即对该平台上的期刊论文、在线图书章节及参考工具书、数据库条目的所有可检索字段 (如题名、著者、著者地址、文摘、关键词等) 进行检索。检索时, 在检索框中直接输入检索词或含有布尔逻辑运算符的检索式, 点击  即可完成检索; ②检索范围选“Publication titles”, 则检索结果为出版物名称中包含该检索词或检索式的文献。例如输入“gene therapy”, 点选“Publication titles”, 则检索出所有出版物名称中含有“gene therapy”的期刊、在线图书等。

(2) 高级检索: 在 Wiley Online Library 主页面选择“Advanced search”, 进入高级检索页面 (图 6-12)。高级检索可进行组合检索及字段限定检索, 还可以对检索年限进行限定, 使检索结果更加精确。检索时, 先在 SEARCH FOR 之下的检索框中输入检索词或含有布尔逻辑运算符的检索式, 检索字段默认为 All Fields, 可以通过下拉菜单选择出版物名称、论文题名、著者、全文、摘要、关键词、资助机构 (Funding Agency)、ISBN、ISSN、论文的 DOI (Article DOI)、参考文献等字段。如果进行多个字段的组合检索, 可以在下拉菜单中选 AND、OR、NOT 逻辑运算符。检索年限 (SPECIFY DATE RANGE) 默认为所有年限 (All Dates), 可以选最近 6 个月、1 个月、2 个月、3 个月, 也可以选一定的年份范围 (Between ...—..., 最早可回溯至 1799 年)。高级检索页面的右侧为检索提示 (Search Tips), 详细给出了 Wiley Online Library 平台的检索规则。

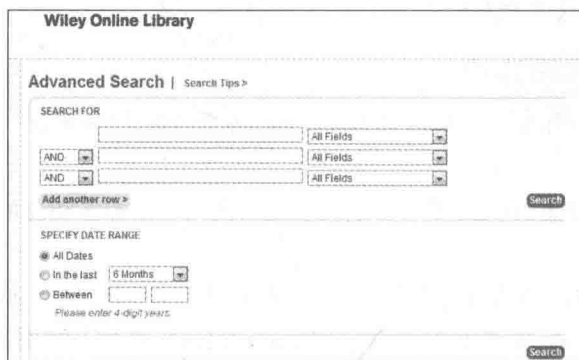


图 6-12 Wiley Online Library 平台的高级检索页面

三、检索结果的处理

检索结果显示页面如图 6-13。首先详细描述检索策略及检索结果的篇数, 其下分别列出在期刊、在线图书等不同产品的检索结果, 给出每篇文献的题录信息, 包括篇名、著者、来源 (刊名及年卷期页码, 如果是 Early View 的文献则给出在线出版的时间)。点击“Abstract”链接进入文摘浏览页面, 点击“HTML”或“PDF”链接可以显示 HTML 格式或 PDF 格式的全文, 点击“References”链接进入参考文献浏览页面。检索结果默认按匹配程度 (Match%) 排序, 点击“Date”可以按出版日期排序。



点击检索结果显示页面右上方的“Edit search”，则返回高级检索页面，可以重新编辑、修改检索策略；点击“Save search”，注册用户可以将检索策略保存在 My Profile 中，方便日后调用、修改或通过电子邮件定期发送检索结果（见图 6-13）。检索结果显示页面右侧的“FILTER LIST”将检索结果进行了期刊、图书、数据库、实验指南的分类。



图 6-13 Wiley Online Library 平台的检索结果显示页面

四、个性化功能与服务

Wiley Online Library 通过 My Profile 建立个性化的服务平台。用户成功注册并激活账号后，在 Wiley Online Library 页面的最上方会出现一个 My Profile 的个性化导航条，导航条右下方会显示 Wiley Online Library 当前在线产品的总数量，可以显示注册用户保存的论文、保存的期刊、保存的检索策略等详细信息（图 6-14）。



图 6-14 Wiley Online Library 平台的 My Profile 页面

1. 远程漫游(Roaming Access) My Profile 页面左侧提供远程漫游的激活和刷新。如果用户的 IP 地址在 Wiley Online Library 授权的 IP 地址范围内, 点击“激活远程漫游(Activate Roaming Access)”, 可实现远程漫游功能。激活漫游功能后 90 天内, 用户在任何地点访问 Wiley Online Library, 只需登录并进入 My Profile, 即可浏览下载全文, 而不受 IP 的限制。90 天后如果仍然需要漫游, 可重新激活此功能。

2. 在 My Profile 主页面, 点击左栏的“Saved Articles”, 显示用户保存的论文题名。

在 My Profile 主页面, 点击左栏的“Saved Searches”, 显示用户保存的检索式。点击“Run Search”直接运行该检索式; 点击“Alert Manager”可设置快讯服务, 即该检索式只要有更新的检索结果, 就通过电子邮件自动发送到用户指定的邮箱。

(孔瑞珍)

第四节 SpringerLink

一、概 述

SpringerLink 是德国施普林格(Springer-Verlag)科技出版集团出版的在线科学、技术和医学(STM)领域学术资源平台。截至 2014 年, 施普林格已出版将近 3000 种科技期刊和 17 万余种科技图书(年出版量近 4000 种), 以及丛书(series)、指南(protocols)、参考作品(reference works)等类型的文献, 学科范围包括自然科学、社会科学、医学以及建筑等各个学科领域。2005 年 Springer 收购 Kluwer 电子期刊后, 资源总量进一步扩大。

1996 年开始推出电子期刊全文数据库检索平台 SpringerLink, 涵盖施普林格出版的所有在线资源, 包括期刊、丛书、图书、参考工具书以及回溯文档, 可进行多语种、跨数据库的信息检索。2002 年 7 月开始, Springer 公司在中国开通了 SpringerLink 服务, 提供简体中文检索界面(图 6-15)。其网址为: <http://www.springer.com.cn>。

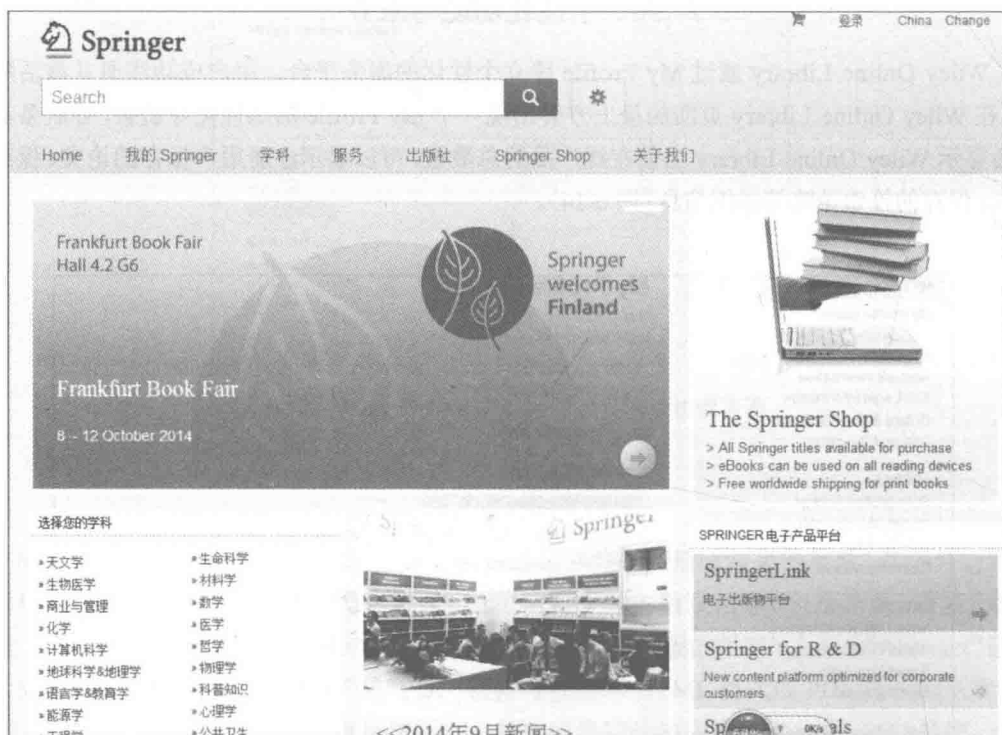


图 6-15 SpringerLink 中文检索主界面



二、检索方法与技巧

SpringerLink 提供的检索方式简单、易用,主要包括浏览和检索两种方式(图 6-16)。同时系统提供灵活实用的限定以缩小检索范围,以便快速获得更为专指的检索结果。

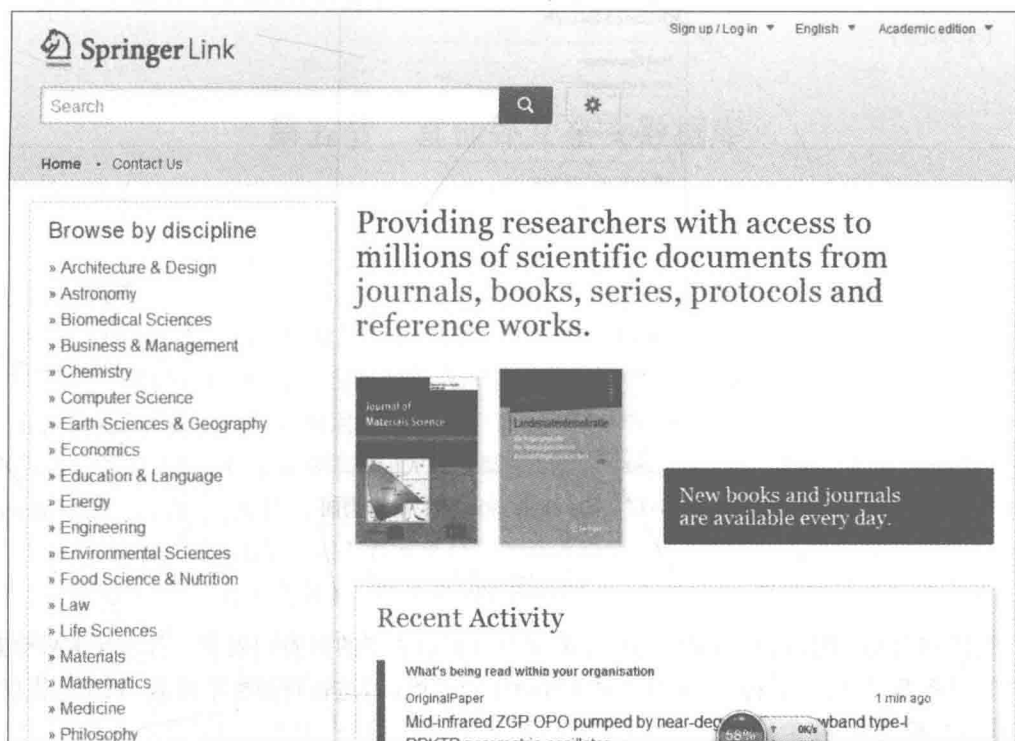


图 6-16 SpringerLink 检索主界面

(一) 浏览

1. 按照学科主题类浏览 SpringerLink 目前包含 24 个学科主题,分别为建筑和设计、天文学、生物医学科学、业务和管理、化学、计算机科学、地球科学与地理、经济学、教育和语言、能源、工程、环境科学、食品科学与营养、法律、生命科学、材料、数学、医学、哲学、物理、心理学、公共卫生、社会科学、统计数据,与医学相关的学科内容集中在“医学”“生物医学科学”和“生命科学”类目中。点击相应的类目即可出现检索结果。

2. 按照资源类型浏览 SpringerLink 按照出版物的形式分成“文献”“图书章节”“参考工具书”“操作指南”等,点击各个类型后即显示该类型的所有书刊名称。比如点击“文献”,即显示 SpringerLink 平台上的所有文献。Protocols 是全球重要的经同行评审的在线实验室指南数据库,收集超过 36 000 条标准化的实验室操作记录,学科领域主要包括生物化学、分子生物学、生物医学等。

(二) 检索

SpringerLink 分为简单检索和高级检索两种方式。

1. 简单检索 是 SpringerLink 的默认检索方式,直接在检索框中输入任意检索词即在全字段中检索。检索词可以是一个词或一组词,也可以是逻辑运算符 AND、OR、NOT 组合成的检索式。

2. 高级检索 用户可以通过 SpringerLink 的高级检索功能准确表达自己的检索需求。点击主页上方的*即可出现 Advanced Search 的按钮,点击此按钮进入高级检索的界面(图 6-17)。其中“with all of the words”表示输入的两个检索词是逻辑与的关系;“with the exact phrase”表示

强制短语检索;“with at least one of the words”表示输入的两个检索词是逻辑或的关系;“without the words”表示输入的两个检索词是逻辑非的关系;“where the title contains”表示限定检索,即检索词出现在篇名中的文献;“where the author / editor is”表示作者检索;“Show documents published”可以设置检索年代。

Advanced Search

Find Resources

with all of the words

with the exact phrase

with at least one of the words

without the words

where the title contains

e.g. "Cassini at Saturn" or Saturn

where the author / editor is

e.g. "H.G. Kennedy" or Ernie Morrison

Show documents published

between and

Include Preview Only content

图 6-17 SpringerLink 高级检索页面

三、检索结果的处理

使用浏览功能时,检索结果为符合相应学科或出版类型的出版物列表。若进行学科检索,系统提供资源类型、子学科、语种等筛选条件供用户使用;若进行资源类型检索,系统则提供学科、语种等筛选条件。

使用检索功能时,检索结果为命中的文献(图 6-18)。检索结果页面左侧将结果按内容类型、学科、二级学科、出版物、语种进行了分类。页面中间显示文献题录,“Sort By”可选择按相

Refine Your Search

Content Type

Article

Discipline

see all

Medicine 819

Biomedical Sciences 379

Life Sciences 181

Psychology 34

Public Health 27

Subdiscipline

see all

Internal 405

Pediatrics 169

Critical Care and Emergency Medicine 160

Biochemistry & Biophysics 118

Anesthesiology 105

Published In

see all

Intensive Care Medicine 96

Diabetologia 35

Critical Care 33

Pediatric Nephrology 27

Current Hypertension Reports 24

Language

English 1,070

German 85

Sort By Relevance Date Published

Page 1 of 58

Article

Defective suppression of the aldosterone biosynthesis during stroke permissive diet in the stroke-prone phenotype of the spontaneously hypertensive rat

Previous studies have shown that short-term high salt intake unmasks blunted plasma aldosterone suppression in stroke-prone spontaneously hypertensive rats (SHRsp). The aim of this study was to evaluate the r...

Iolanda Enea, Paola De Paolis, Antonio Porcellini... in *Basic Research in Cardiology* (2000)

» Download PDF (160 KB)

Article

Oxidative stress plays a permissive role in $\alpha 2$ -adrenoceptor-mediated events in immortalized SHR proximal tubular epithelial cells

The present study evaluated the role of oxidative stress on $\alpha 2$ -adrenoceptor-mediated events (Cl⁻/HCO³⁻ exchanger activity and cAMP accumulation) in immortal...

Sónia Simão, Sónia Fraga, Pedro A. Jose... in *Molecular and Cellular Biochemistry* (2008)

» Download PDF (467 KB) » View Article

Article

Effects of rapid permissive hypercapnia on hemodynamics, gas exchange, and oxygen transport and consumption during mechanical ventilation for the acute respiratory distress syndrome

To measure the effects of rapid permissive hypercapnia on hemodynamics and gas exchange in patients with acute respiratory distress syndrome (ARDS).

J.-B. Thorens, P. Joliet, M. Ritz, J.-C. Chevrolet in *Intensive Care Medicine* (1996)

» Download PDF (573 KB)

Article

图 6-18 SpringerLink 结果显示页面

关度、出版时间进行排序。“Date Published”可将结果按时间进一步筛选。题录下方“Download PDF”和“View Article”分别为 PDF 和 HTML 格式的全文链接。

四、个性化功能与服务

SpringerLink 注册登录后,可显示最近的检索、特色期刊和特色图书,并可保存检索策略。

(孔瑞珍)

第五节 其他外文全文数据库

一、ScienceDirect

(一) 概述

ScienceDirect 是荷兰爱思唯尔 (Elsevier) 出版公司在网上施行的文献信息检索系统。爱思唯尔公司是一家经营科学、技术和医学信息产品及出版服务的世界著名出版公司,隶属于 Reed Elsevier 集团,总部设在荷兰的阿姆斯特丹。通过与全球的科技与医学机构的合作,Elsevier 公司每年出版近 2500 种学术期刊和 1900 种新书,以及一系列的电子产品,如 ScienceDirect、Scopus、MD Consult、在线参考书目等。

Elsevier 公司出版的期刊、图书是业界公认的高质量学术出版物,其出版的期刊大多数为同行评审的核心期刊,被许多著名的二次文献数据库所收录,其中约 1400 种期刊被 SCI 收录。期刊内容涉及 24 个学科,包括数学、物理、生命科学、化学、计算机科学、临床医学、环境科学、材料科学、航空航天、工程与能源技术、地球科学、天文学及经济、商业管理和社会科学等。

Elsevier 公司推出了 ScienceDirect 全文数据库,包括期刊全文、单行本电子书、参考工具书、手册以及图书系列等。2014 年 ScienceDirect 提供的文献量为,期刊近 2500 种,图书近 26 000 册。目前,Elsevier 公司通过其主站点,即 ScienceDirect Online (SDOL) 提供服务(图 6-19)。SDOL 检索平台(<http://www.sciencedirect.com>)在内容更新速度和范围、检索模式、检索技巧和个性化服务等方面都有很大优势。

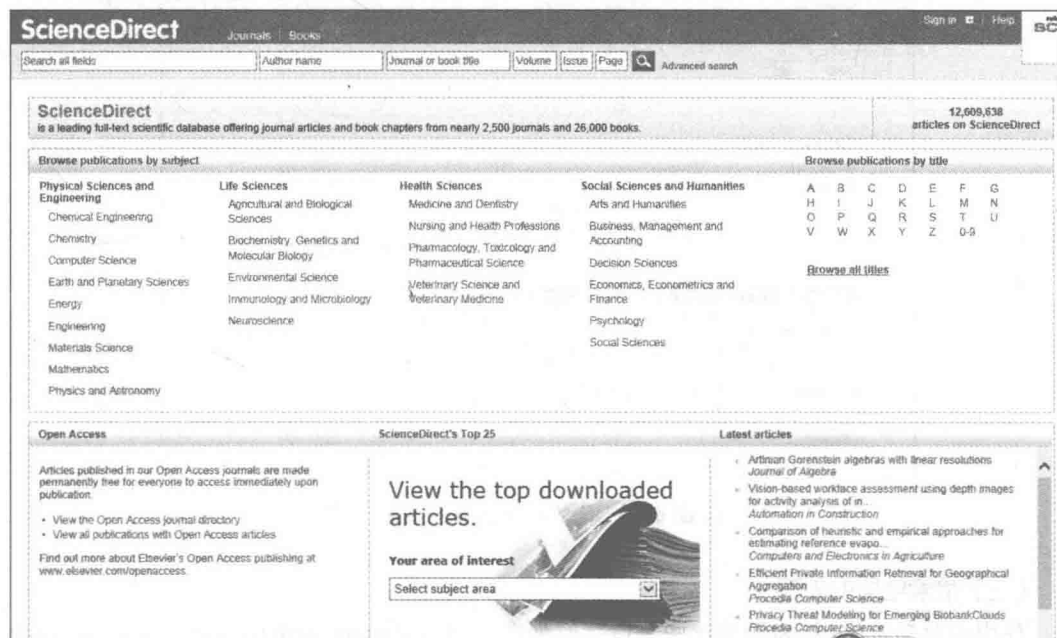


图 6-19 SDOL 数据库主页

(二) 检索方法

SDOL 数据库主要提供浏览和检索两种文献查询方法。其中,检索又包括快速检索、高级检索两种检索方式。

1. 浏览

(1) 按学科主题浏览(Browse publications by subject): SDOL 数据库主页上列出理科与工程学(Physical Sciences and Engineering)、生命科学(Life Sciences)、健康科学(Health Sciences)以及社会学与人文学科(Social Sciences and Humanities)4个大的一级类目,点击类目名称即可浏览该主题的书刊信息,书刊信息按字顺排列。

(2) 按书刊字顺浏览(Browse publications by title): SDOL 数据库主页上将所有期刊按刊名字母顺序排列,点击相应的字母即可进入该字母开头的书刊信息。

2. 检索 SDOL 数据库的检索功能规定了多种检索运算符和检索规则,包括:①布尔逻辑运算符逻辑“与”AND、逻辑“或”OR、逻辑“非”NOT;②截词符:“*”、“?”;③位置算符 W/n 或 PRE/n;④用双引号表达的强制短语检索等。

(1) 快速检索: SDOL 数据库所有界面上方都设有快速检索区,可对文章或期刊进行快速查找。可供选择的检索字段包括所有字段、著者姓名、期刊/图书名称、卷、期和页码等。检索时可选择其中一项或几项内容进行检索,不同字段之间的关系为 AND。如果只用刊名字段检索,检索结果将为刊名包含该检索词的所有期刊列表。其他字段检索将直接得到期刊论文信息。

(2) 高级检索: 点击数据库主页上的“Advanced search”链接即可进入高级检索界面(图 6-20)。高级检索提供两个检索输入框,可以输入单词、词组或布尔逻辑检索表达式。输入框后的字段限制选项默认为 All Fields,也可进行特定字段限制选择。高级检索同时还提供文献来源、学科主题、文献类型、出版时间等限定条件供选择以获得更精确的检索结果。

图 6-20 SDOL 高级检索界面

(三) 检索结果的处理

使用浏览功能时,检索到的出版物以字顺列表(Titles starting with)。页面左侧提供“Filter by subject”,可将检索结果按照学科分类浏览;页面右侧则提供了“All publications”(书刊筛选)、



“All access types”(获取方式筛选)。

使用检索功能时,首先在检索结果页面上方显示命中结果数量、检索式。命中记录信息包括论文题名、刊名、卷期、出版日期、页码、作者等。页面左侧可按年代、期刊、主题、内容类型等不同的方式筛选检索结果。右侧图标亮色表示用户可以查看该论文的全文,右侧图标

(四) 个性化服务

SDOL 提供了提醒、My Favorites 等功能。注册登录后,用户的检索策略可以设置为“检索提醒”;对感兴趣的期刊或以设为“My Favorites”,方便以后阅读和检索。

二、CINAHL PLUS with Full Text

(一) 概况

护理及相关健康文献累积索引(Cumulative Index to Nursing and Allied Health Literature, CINAHL)是 EBSCO 公司出版的专门面向护理及相关专业人员的数据库,内容源于全球英文护理专业期刊、美国护理协会、国际护理联盟组织、护理卫生科学联盟组织等权威机构,涵盖护理学、心理学、行为科学、替代医学、营养学和康复等护理学与综合保健相关学科。美国 EBSCO 公司创建于 1943 年,是全球最早提供数据库在线检索的公司之一,目前所用的检索系统为 EBSCOhost。

CINAHL PLUS with Full Text 是 CINAHL 的全文版,文献最早回溯至 1937 年,收录了包括 770 多种期刊和 270 多种图书在内的全文资源,以及多种会议论文、护理学学位论文、操作案例等,还新增了 140 种疾病的循证护理说明和 180 种快速教学全文,为护理人员自学或患者健康教育提供了参考资料以及疾病护理相关的最新证据。

(二) 检索方法

CINAHL PLUS with Full Text 在 EBSCOhost 平台上运行,提供基本检索、高级检索、主题检索、出版物检索等检索方式,点击系统最上方工具条中的检索方式链接即可切换(图 6-21)。

1. 高级检索 CINAHL 的默认检索方式。高级检索提供多个字段的逻辑组配,可根据需要选定字段后,在检索框中输入检索词,再对 AND、OR 或 NOT 等逻辑运算符进行选择后,在检索选项中首先对布尔逻辑检索、查找全部检索词、查找任何检索词、智能检索等检索模式和扩展条件进行选择,再对文摘、作者、语种、研究条件等进行限制后,即可完成检索。

正在检索 CINAHL Plus with Full Text | 选择数据库

选择一个字段(可选) | 检索 | 清除

AND | 选择一个字段(可选)

OR | 选择一个字段(可选)

NOT | 选择一个字段(可选)

基本检索 | 高级检索 | 检索历史记录

检索选项

检索模式和扩展条件

检索模式 ?

☒ 布尔逻辑/逻辑

☐ 查找全部检索词

☐ 查找任何检索词

☐ 智能检索/语义

应用相关字词

☐ 也检索全文文章

限制结果

全文 ☒

有文摘 ☐

作者 ☐

高质 ☐

相关全文 ☐

有参考 ☐

出版日期

月 年 月 年

出版物

同行评审 ☐

不包括 Pre-CINAHL ☐

图 6-21 CINAHL PLUS with Full Text 主页

2. 基本检索 方式较为简单,用户只需在检索框内输入要检索的单词、词组或检索表达式后,在页面下方的检索选项部分对检索模式和扩展条件、检索结果进行限定即可(图 6-22)。



图 6-22 CINAHL PLUS with Full Text 基本检索页面

3. 主题检索 CINAHL 的主题检索类似 PubMed, 其主题词、副主题词主要从 MeSH 中抽取, 偏重疾病、药物、解剖、生理等方面, 并新增了护理及相关学科的专业词汇, 共计 14 000 余条。用户可根据需要组配检索(图 6-23)。

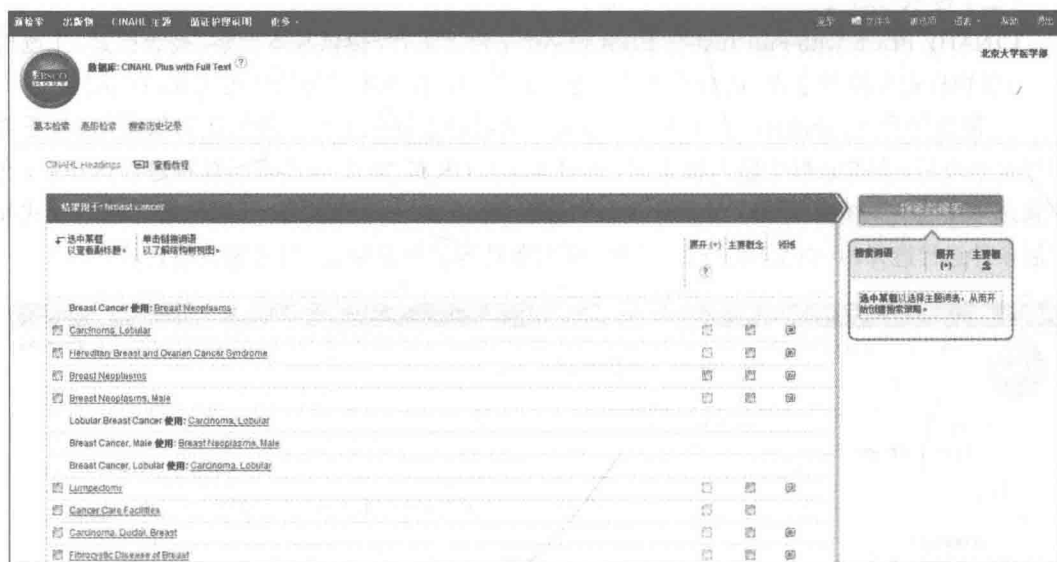


图 6-23 CINAHL PLUS with Full Text 主题检索页面

4. 出版物检索 通过出版物检索页面,可浏览系统收录的期刊信息。在检索词输入框内输入关键词,选择“按字母顺序”“按主题和说明”或“匹配任意关键字”排序,点击“浏览”后,勾选选中的期刊并单击“添加”按钮,即可查看出版物详细资料(图 6-24)。

5. 循证护理说明检索 该检索方式主要对护理实践中的重要主题如具体疾病进行检索,检索结果以 PDF 文档的形式呈现,内容为与所检索的题目或主题相关的最新循证证据、统计数据、研究和参考文献等。检索时直接在检索词输入框内输入关键词,选择按字母或相关性的排序顺



序后,点击“浏览”按钮,在结果页面勾选相关的循证护理主题后,即得到结果。结果文档的内容统一由 What We Know 和 What We Can Do 两部分组成。

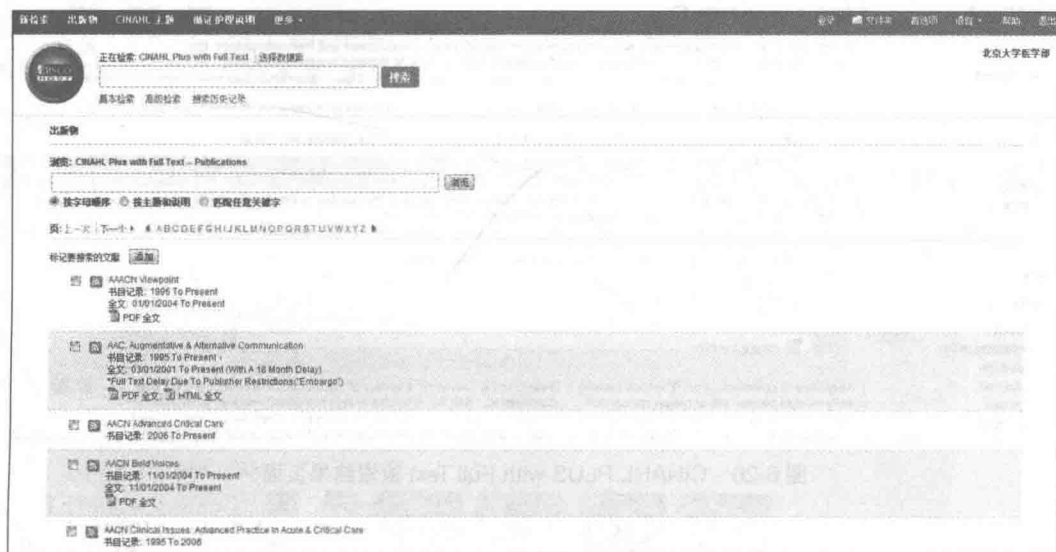


图 6-24 CINAHL PLUS with Full Text 出版物检索页面

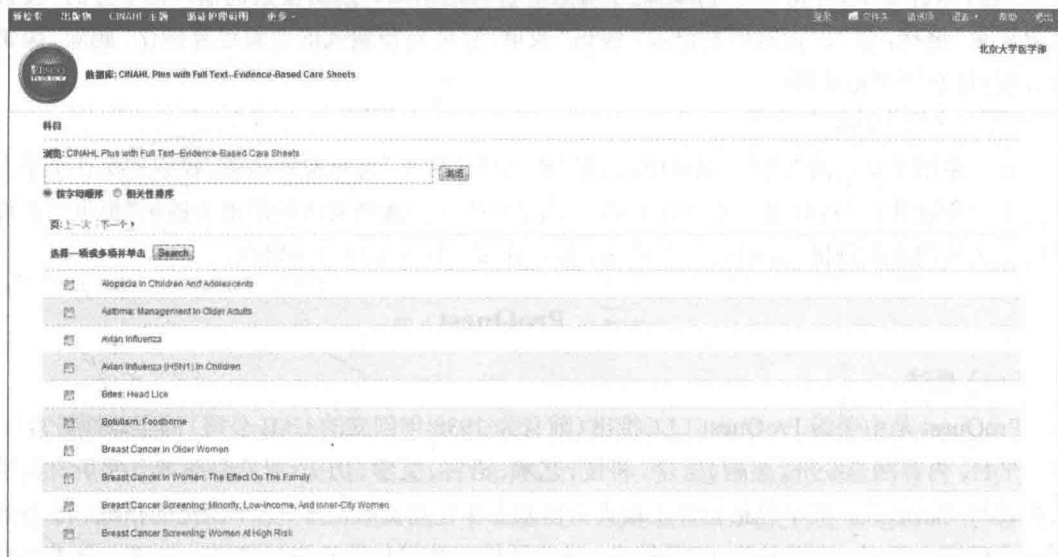


图 6-25 CINAHL PLUS with Full Text 循证护理说明检索页面

6. 快速教学检索 点击数据库最上方工具条中的“更多”下拉菜单,选择“快速教学”即可进入检索界面,直接输入检索词即可检索。快速教学是依据护理工作的流程组织的相关信息,如疾病的病因、症状体征、诊断评估、(护理干预涉及的)治疗目标、患者及家属宣教等,总结临床护理的概况,以供用户自学或健康教育参考(图 6-25)。

此外,系统还提供参考文献、图像和索引的检索,在系统最上方工具条中的“更多”下拉菜单中可以进行选择。

(三) 检索结果的处理

CINAHL PLUS with Full Text 的检索结果页面分左右两栏,左侧为结果筛选区,右侧为结果显示区(图 6-26)。

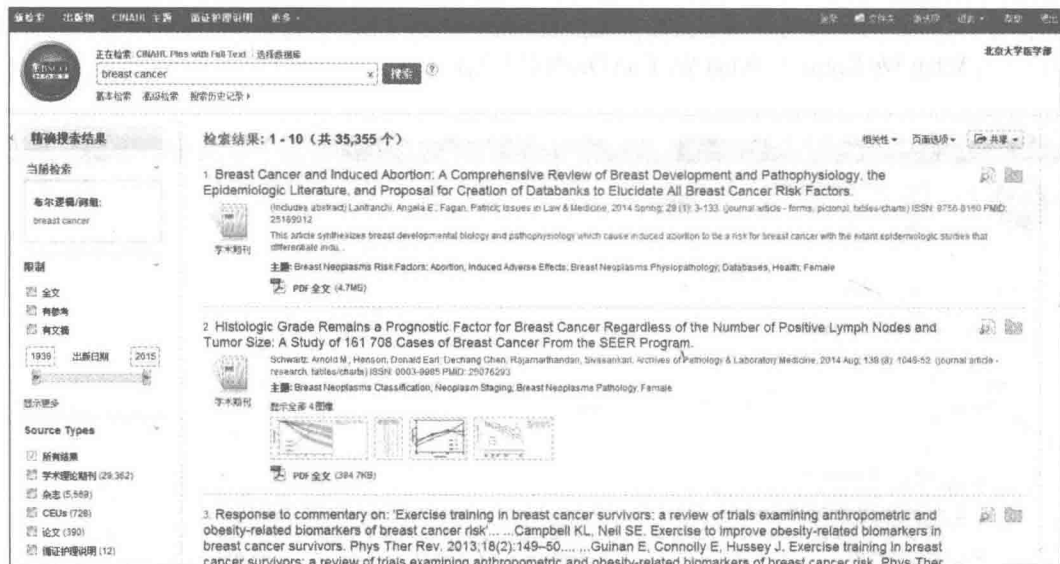


图 6-26 CINAHL PLUS with Full Text 检索结果页面

系统默认按相关性对结果进行排序,也可按日期、来源、作者等对检索结果进行排序。可以重新对结果格式、图像、每页的结果等对页面重新设置,也可以通过限定出版日期、文献类型、主题、来源文献、年龄、性别和地理学等进一步筛选结果。点击检索词输入框下方的“搜索历史记录”链接,进入“搜索历史记录/快讯”表单,完成对检索式的逻辑运算操作、删除、编辑或打印、保存历史记录等。

(四) 个性化功能与服务

在检索结果页面的“共享”菜单中,选择“添加至文件夹”可将检索结果、检索式保存至个人文件夹;“创建快讯”可以电子邮件或 RSS 形式定时推送检索结果;“使用永久链接”提供当前检索结果页面的永久链接,点击其下方书签图标可在多个网站上分享该链接。

三、ProQuest

(一) 概述

ProQuest 是由美国 ProQuest LLC 推出(前身为 1938 年创立的 UMI 公司)的全球性检索与研究平台,内容涵盖经济、金融、医学、科技、艺术、语言、文学、历史、社会科学等几乎所有的学科领域和文献形式。其中包括世界上最大的博硕士学位论文数据库、三个世纪的报纸、45 万余册学术研究电子书、绝版书籍、记录档案、以及原始的数据与原始资料等学。该平台包含多个学术性数据库,其中使用较多的数据库如世界最大的联机检索系统 Dialog、硕博士论文数据库 PQDT、电子图书数据库 ebrary 和 EBL 以及发现系统 Summon 等。平台中的常用医学数据库有:

1. ProQuest Health & Medical Complete (PHMC) 该库收录 3100 多种出版物,其中 2600 多种有全文,范围涵盖全部重要的临床和健康护理领域。
2. ProQuest Dissertations & Theses (PQDT) 收录 290 多万条记录,是全球收录最广泛的博硕论文文摘数据库。
3. ProQuest Nursing & Allied Health Source (PNAHS) 收录护理学、保健医学、替代医学等学科的 1000 多条期刊文献记录、15 700 多条护理学学位论文记录、电子图书、影像等。
4. ProQuest Research Library (PRL, 原 ARL) 综合性学术研究数据库。该库提供 6300 多种出版物,其中包括 4700 多种全文出版物,最早可回溯至 1933 年。其主题范围涵盖 150 个

核心学术主题,为广泛的学术领域提供一站式信息服务。

(二) 检索方法

ProQuest 提供基本检索、高级检索、出版物检索和浏览四种检索方式。在执行检索命令前,系统允许对检索的数据库进行限定,可以选择一个或多个数据库,还可以选择检索的学科范围,如健康与医学、科学与技术等(图 6-27)。



图 6-27 ProQuest 主页

1. 基本检索 是 ProQuest 默认的检索方式,用户只需直接在检索词输入框中输入单词或词组即可,检索时系统自动给出检索建议供选择。用户也可以通过全文文献、同行评审和学术期刊三个选项对检索的期刊范围进行进一步的限定。基本检索方式支持逻辑运算符 AND(逻辑与)和 OR(逻辑或)以缩小或扩大检索范围,输入的检索词之间的关系默认为 AND。

2. 高级检索 界面提供多个检索词输入框和字段供选择,并选择出版日期、出版物类型、文档类型、语言、排序选项、每页显示条目等多种限制选项。ProQuest 支持 AND、OR、NOT、NEAR/N、PRE/N 等多种运算符,用户也可以选择使用字段和运算符书写命令行的形式进行检索(图 6-28)。此外,高级检索方式支持对报告、数据与研究报告、图和表的查找,并可以使用高级检索中的“以引文查找全文”选项,使用出处信息查找文献。

3. 出版物检索 ProQuest 支持在所选数据库中检索和浏览全文文献出版物,可在检索词输入框中输入期刊标题的全称、部分内容进行检索。系统支持以“*”结尾的截词符检索,并可限定检索词在期刊的标题中、标题开头、摘要或主题等位置(图 6-29)。还可以使用出版物类型、出版物主题、语言、出版商、数据库等对检索范围进行限定。

4. 浏览 用户可以通过该检索方式浏览 ProQuest 各数据库的主题和特色内容。提供按主题和地点浏览两种方式(图 6-30)。



图 6-28 ProQuest 高级检索界面



图 6-29 ProQuest 出版物检索界面



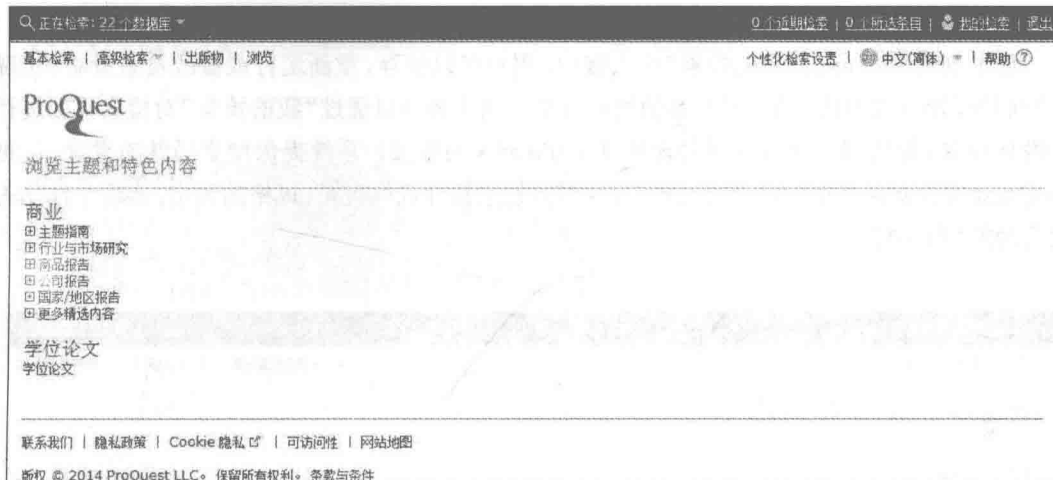


图 6-30 ProQuest 浏览界面

(三) 检索结果的处理

在检索结果页面中,系统在检索框下方列出相关的检索主题备选,用户还可以根据右侧列出的全文文献、同行评审、学术期刊、出版物类型、出版物名称、文献类型、主题、分类等多个分类选项进一步缩小检索结果范围。检索结果默认按相关性排序,检索词在每条记录中均高亮显示,点击预览按钮可以查看到更为详细的记录信息,或根据需要查看具体摘要、下载全文、选择以其他途径请求获取文献或订购文献(图 6-31)。如需保存或打印文献,可以选中记录前的复选框标记文献后,点击检索界面上方的打印、导出/保存等相关按钮。



图 6-31 ProQuest 检索结果界面

(四) 个性化功能与服务

通过创建 ProQuest“我的检索”个人账户,用户可以保存、重新运行或修改检索策略,实现定题服务,保存需要的文献,并与其他用户共享。用户还可以通过“我的检索”对检索界面进行个性化设置,并实现与参考文献管理软件 RefWorks 的连接。系统提供检索插件的设置,可将 ProQuest 的检索框添加到自己的网页中,同时对检索插件进行布局、风格的设定,实现个性化的检索功能(图 6-32)。



图 6-32 ProQuest 个性化功能与服务界面

四、Karger

(一) 概述

Karger——Medical and Scientific Publications 是 S. Karger 出版社推出的生物医学文献检索数据库。

S. Karger 出版社是瑞士一家独资家族出版社,成立于 1890 年。Karger 出版社以医学为主,是世界上为数不多的完全专注于生物医学出版领域的出版社,也是世界上享有盛誉的医学出版社之一。Karger 每年出版约 90 余种期刊(包括 17 种在线开放存取期刊、附属期刊以及病案报告,绝大部分被 SCI 收录)、40 种丛书和每年为数不少的专著。主要为英语出版物,涵盖了传统医学和现代医学的最新发展。

Karger 只发表、出版社经过同行评审后的稿件,因而,出版社具有高质量的制作和内容。Karger 绝大多数的图书和期刊都有电子版,而且许多期刊都在线上优先发行。

(二) 检索方法

Karger 主要提供对出版物或文献的浏览和检索功能。检索功能又有基本检索和高级检索两个入口(图 6-33)。

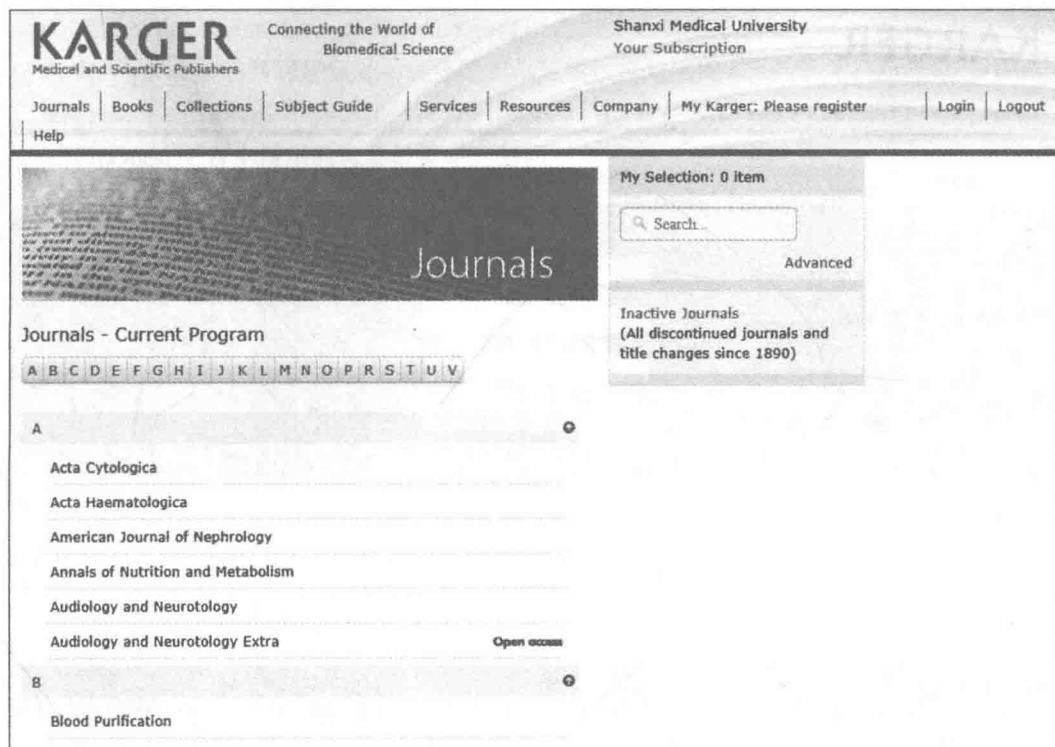


图 6-33 Karger 检索主页

1. 浏览

(1) 按字顺浏览: Karger 可通过“Journal”和“Book”两个入口分别对期刊和图书按字顺进行浏览。在图书浏览页中,尚有“Book News”“Book Series”“Non-Serial Books”“Complete Title Index”“Multimedia”等种类可供选择,分别检索“图书新闻”“丛书”“非丛书”“综合索引”“多媒体”方面的图书信息。

(2) 学科浏览: Karger 提供了“Subject Guide”(学科指南)入口,向用户提供通过专业医学学科进行出版物浏览的功能。点击主页的“Subject Guide”,系统列出医学学科列表,点击特定学科名称,即可获得属于该学科的 Karger 出版物。

2. 基本检索 Karger 的主页上提供了基本检索区。检索方法十分简单,在检索框内输入检索内容(词语间的默认运算关系为逻辑与),回车即可完成检索。对于精确短语支持双引号的强制检索。

3. 高级检索 点击 Karger 主页中基本检索区下方的“Advanced”按钮可进入高级检索页面(图 6-34)。利用高级检索功能,用户能够准确表达自己的检索需求。

页面的 Find results 之下,“with all of the words”表示输入的检索词均需出现在检索结果中;“with the exact phrase”表示输入内容以精确短语形式检索,不能拆分;“with at least one of the words”表示输入的检索词只需有一个出现在检索结果中即可;“without the words”表示去除该检索词。

高级检索中也提供了相应的限制功能,包括:位置(Occurrences,标题或任意位置)、排序方式(Sort,相关度或日期)、出版时间等。

(三) 检索结果处理

Karger 的浏览功能中,若按学科指南进行浏览,则系统提供属于某一学科的 Karger 出版列表,点击特定出版物,可查看该出版物的详细信息:若为期刊,则列出该刊的历史卷期刊供用户选择、查看、输出;若为图书,则直接列出该书的详细信息。

图 6-34 Karger 高级检索页面

使用 Karger 的基本检索和高级检索功能,均可得到特定的文献。系统直接提供该文献的题录、摘要和全文信息。同时,系统还设置了文献工具功能,用于链接该文献的进一步操作或与其他系统或数据库的相关信息。

(四) 个性化服务

Karger 向用户提供的个性化服务主要是提醒功能。用户注册登录后,可按自己的要求接受系统提供的提醒服务。

本章小结

本章主要介绍了 ClinicalKey、Journals@Ovid Full Text、Wiley Online Library、SpringerLink、ScienceDirect 等目前常用的外文电子文献全文数据库的基本情况、检索方法与技巧、检索结果的处理方法,以及个性化服务内容与方式等,对于用户进行外文文献检索、获取原始文献提供有力帮助。

(谢志耘 孔瑞珍)

思考题

1. ClinicalKey 包含了哪些类型的全文? 各有什么特点?
2. ClinicalKey 的词语检索有什么特殊功能?
3. OvidSP 的高级检索(Advanced Search)怎样使用?
4. OvidSP 平台中不同的数据库系统使用的“Search Tools”控词词表相同吗?
5. 简述 Wiley Online Library 的检索规则。

6. 应用 Wiley Online Library 浏览出版物时可使用哪些浏览方式?
7. SpringerLink 的高级检索是如何体现布尔逻辑运算规则的?
8. ScienceDirect 提供哪些个性化服务?
9. ProQuest 平台中常用的医学数据库有哪些?
10. Karger 的检索结果显示页面中, 文献工具(Article Tools)有什么功能?

第七章 引文检索

作者撰写论文时,为了充实文章内容,增添理论色彩,增强论辩的力量,往往会引用其他文献。这表明一项学术研究成果通常是在前人基础上完成的,它体现着学术研究的继承性和发展性。引文检索就是通过检索这些存在互相引用关系的文献,获取相关主题的最新文献,揭示学者之间的学术交流,对研究人员的科研绩效和期刊质量进行评价。

第一节 引文检索概述

一、基本概念

引文,又称被引用文献(cited paper),指列于文献末尾或脚注的参考文献(reference)。列有参考文献的文献称为引用文献(citing paper)又称引证文献。引用文献与被引用文献是相对的。A文献对于它的参考文献来说就是引用文献,但当A文献被别人引用后就成了那篇文献的被引用文献。当作者引用自己发表的文献时称为作者自引,同一期刊上文献的互相引用是期刊自引,二者统称自引(self-citation)。

引文检索(citation search)是指以被引用文献为检索起点来检索文献的被引用情况。引文检索中最常见的检索词有被引作者、被引刊名、被引年份和被引文献标题词等。

引文检索时需要使用专门的检索工具。常见的引文数据库有Web of Science核心合集、中国科学引文数据库、中国引文数据库等。

二、引文检索的作用

引文检索在学术交流和科研评价中的作用日显突出,其作用有以下几个方面。

1. 获得同一主题相关的新文献 由于被引文献和引用文献在内容上或多或少有关联,因此通过一位知名学者或一篇较有质量的文献进行引文检索,常常可以获得一系列主题相关、内容上有所继承和发展的新文献,当然也可能检索到个别持商榷反对意见的文献。

2. 用于评估学术论文的影响力 某篇文献发表后,它的被引用次数(times cited)可能会从零逐渐增多,尤其是高质量的原创文献。论文被其他文献正面引用,是其学术观点和研究成果被他人参考借鉴认可的例证。被引用频次越高,表示论文的影响力越大。

3. 用于评估研究人员的学术水平 文献质量与文献被引次数成正比已被广泛认同。在晋升职称和引进人才之前,目前常见的做法是请具有检索资质的图书馆出具查收查引报告,将被他人引用频次作为被考察者学术水平高低的依据之一。

4. 用于评估机构或国家的科研实力 对一所科研机构、一所大学乃至一个国家而言,文献被引总频次在一定程度上能反映该机构该国家科研的总体实力。文献被引总频次主要取决于文献发表量和文献本身的学术质量。这类评估可通过专门的课题研究来完成,也可通过引文数据库的副产品Essential Science Indicator(基本科学指标,ESI)等来查询。

5. 用于评价学术期刊的质量 学术期刊质量的评价因素很多。目前国际上统一采用的计量指标有期刊的影响因子、即年指数和被引总频次等,这些数据都可从引文数据库的副产品

Journal Citation Reports(期刊引用报告, JCR)中获得。

6. 为学科发展研究提供计量数据 在观察学科之间的渗透交叉、测定学科文献老化速度、研究文献引证规律、进行文献计量研究等方面, 引文数据库及其副产品“期刊引用报告”、“基本科学指标”等都是不可或缺的工具。

引文检索在评价科研人员学术水平和期刊学术质量方面所起的作用毋庸置疑, 但也存在局限性。例如有争议的论文易被引用, 好友师生之间的过度引用会使文献的被引频次虚高; Web of Science 核心合集收录的期刊大多数质量较高, 但也有影响因子小于 0.01 的; 有一定学术质量, 但未被 Web of Science 核心合集收录的期刊, 尤其是非英语期刊, 在引文检索中易受冷落; 部分期刊过度自引影响了该刊影响因子的真实性。因此, 引文检索和引文数据可作为科研绩效评价的依据, 但不能作为唯一的依据, 这已成为学术界和科研管理部门的共识。

第二节 Web of Science 核心合集

一、概 述

Web of Science(WOS, 曾名 Web of Knowledge)是 Thomson Reuters 公司的学术信息资源整合平台。该平台整合的书目数据库有 Web of Science 核心合集、Current Contents Connect、Derwent Innovations Index、BIOSIS Previews、MEDLINE、SciELO Citation Index、中国科学引文数据库等, 其中 Web of Science 核心合集、Derwent Innovations Index、中国科学引文数据库有引文检索功能。分析工具有 Journal Citation Reports(期刊引用报告, JCR)和 Essential Science Indicators。另外提供一部分 ISI 外部资源链接。

WOS 功能齐全, 具有所有数据库检索、单库检索、引文检索、定题快讯服务、引文跟踪服务、创建引文报告、检索结果分析、检索结果提炼、期刊影响因子查询、期刊定制和个人文献管理等功能。

WOS 中影响力最大的数据库是 Web of Science 核心合集, 由 5 个引文数据库和 2 个化学数据库组成。收录来自各个研究领域的数千种学术期刊及会议录上的文献信息。

1. Science Citation Index Expanded(科学引文索引扩展版, SCI-Expanded) 收录 8300 多种科技期刊, 最早可回溯至 1900 年。涵盖学科有: 农业、天文学、生物化学、生物学、生物工艺学、化学、计算机科学、材料学、数学、医学、精神病学、肿瘤学、药理学、儿科学、物理学、植物学、精神病学、外科学、兽医学和动物学等。

2. Social Sciences Citation Index(社会科学引文索引, SSCI) 收录 2900 多种社会科学期刊和选择性收录期刊 3500 多种, 最早可回溯至 1900 年。涵盖学科有: 人类学、历史、工业关系、情报学和图书馆学、法学、语言学、哲学、心理学、精神病学、政治学、公共卫生、社会问题、社会工作、社会学、药物滥用、城市研究和妇女研究等。

3. Arts & Humanities Citation Index(艺术与人文科学引文索引 A&HCI) 收录 1600 多种艺术与人文科学期刊和选择性收录期刊 6000 多种, 最早可回溯至 1975 年。涵盖学科有: 考古学、建筑学、艺术、亚洲研究、古典作品、舞蹈、民间传说、历史、语言、语言学、文学评论、文学、音乐、哲学、诗歌、广播影视、宗教和戏剧等。

4. Conference Proceedings Citation Index-Science(科学会议录引文索引 CPCI-S) 收录 1990 年至今的自然科学方面的会议文献。

5. Conference Proceedings Citation Index-Social Sciences & Humanities(社会科学会议录引文索引, CPCI-SSH) 收录 1990 年至今的社会科学与人文科学方面的会议文献。

6. Index Chemicus (IC) 收录 1993 年以来国际一流期刊上报道的新的有机化合物的化学结构与评论数据, 其中许多记录展示了从最初的原料到最终产品的整个化学反应过程。IC 是揭示生物活性化合物和天然产品最新信息的重要信息源。

7. Current Chemical Reactions (CCR) 包含摘自知名期刊和 36 家专利授予机构的单步骤或多步骤新合成方法。所有方法均带有总体反应流程, 且每个反应步骤都配有详细和准确的图形。该数据库同时收录 14 万条来自法国工业产权局 (INPI) 的化学反应记录, 最早的数据可回溯至 1840 年。

二、检索规则

Web of Science 核心合集主要包含以下一些检索规则。

1. 支持 AND、OR、NOT 布尔算符 运算符大小写皆可。系统默认空格为 AND 逻辑运算, 同时遵循 NOT>AND>OR 的运算顺序。

2. 词组检索时可使用双引号 表示检出文献中含有于引号内完全相同的检索词。

3. 提供 NEAR/X 位置限定运算符 表示由该运算符连接的检索词之间相隔不大于 X 的单词。如果只使用 NEAR 而不使用“/x”, 则系统将查找其中的检索词由 NEAR 连接且彼此相隔不大于 15 个单词的记录(详见第一章第三节)。

4. 提供 SAME 位置限定算符 使用 SAME 将得到检索词出现在同一地址字段的记录。您需要使用括号将检索词括起来。例如: AD=(McGill Univ SAME Canada) 得到在“地址”字段中出现 McGill University 以及 Canada 的记录。当在其他字段(如“主题”和“标题”)中使用时, SAME 与 AND 的作用就相同。

5. 提供星号*、问号? 和美元符号\$ 三种通配符 “*”替代多个字符, 一般跟在单词后面, 表示检索相同词根不同后缀的检索词。

三、检索方法和技巧

Web of Science 核心合集提供基本检索、作者检索、被引参考文献检索、化学结构检索和高级检索等多种检索方法。其中除“被引参考文献检索”这种特定文献被引情况检索方法外, 其余检索方法都属于常规的检索方法。

1. 基本检索 是默认的检索页面, 提供的检索字段有主题、标题、作者、团体作者、出版物名称、出版年、地址、语种、文献类型等。

【例 1】查 2008—2014 年有关“禽流感”的英文综述文献。在图 7-1 的第一个检索框内输入 "avian influenza" or "bird flu", 检索字段选标题, 点击“添加另一字段”, 第二个检索框后检索字段选语种, 提问框内选 English, 第三个检索框后检索字段选文献类型, 提问框内选 Review, 运算符都选 AND, 时间跨度从 2008—2014, 点击“检索”即可。

主题和标题是最常用的检索字段。由于 Web of Science 核心合集没有主题词, 用主题和标题检索时要考虑同义词情况。本例中的 avian influenza 和 bird flu 是同义词, 都表示禽流感, 所以用运算符“or”连接。为了查全, 还可以用 H5N1、H7N9 等特定禽流感病毒作为检索词参与检索, 用运算符“or”连接。

2. 作者检索 是一种作者甄别检索, 它通过对研究领域的筛选, 准确获得该作者的文献。

【例 2】查找美国贝勒研究所(BAYLOR Research Institute)的 Jacques Banchemreau 教授发表的文献。

检索步骤是: 进入作者检索界面, 在姓(必填)框内输入 Banchemreau, 在首字母框内输入名首字母 J, 点击“选择研究领域”; 研究领域选 LIFE SCIENCES BIOMEDICINE, 点击“选择组

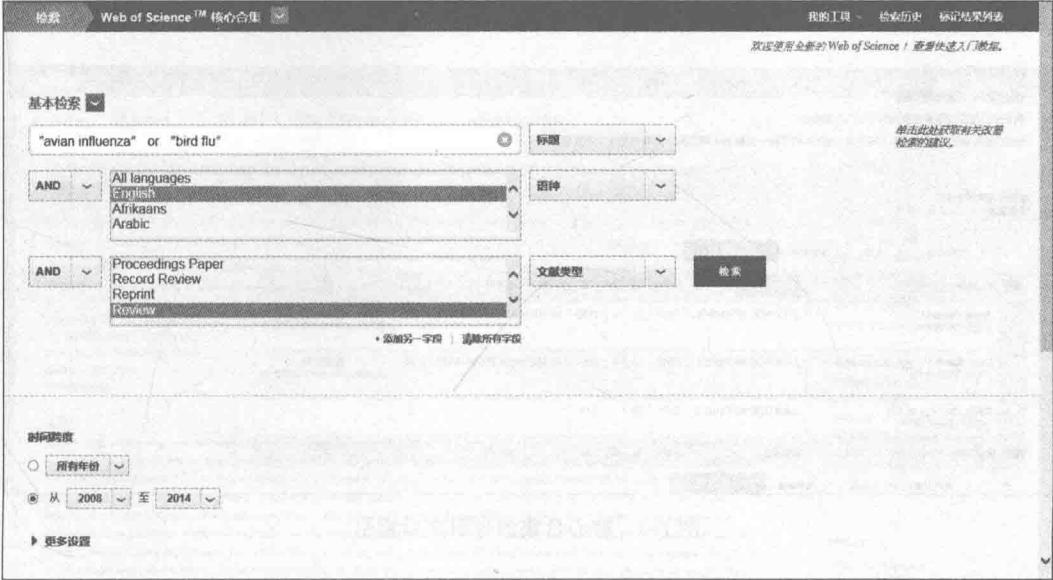


图 7-1 核心合集的基本检索页面

织”；选 BAYLOR RESEARCH INSTITUTE，点击“完成检索”；得到 Jacques Banchereau 教授的 294 篇文献。

3. 被引参考文献检索 即通过被引用文献获得引用文献。检索字段有被引作者、被引著作（即被引文献所在的书刊名）、被引年份、被引卷、被引期、被引页、被引标题。可选择单一字段进行检索，也可以同时进行多字段检索。多字段检索时，字段之间只能用布尔逻辑算符 AND 组配。

【例 3】检索 Gerald B. Appel 2009 年在 *Journal of the American Society of Nephrology* 第 5 期上发表的文献 *Mycophenolate Mofetil versus Cyclophosphamide for Induction Treatment of Lupus Nephritis* 被人引用情况。

检索步骤：在被引参考文献检索页面（图 7-2）的 3 个检索框内分别输入 appel gb、J AM SOC NEPHROL（刊名缩写）、2009，点击检索后，得到被引文献索引（图 7-3）。

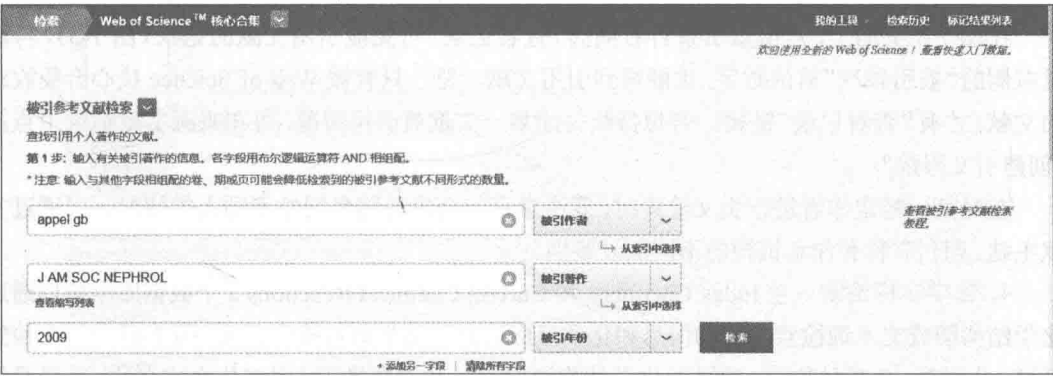


图 7-2 核心合集的被引参考文献检索页面

在图 7-3 中，被引文献索引条目从左至右分别表示：被引作者、被引著作、出版年、卷、期、页、标识符、施引文献数（引用文献数）、查看记录。勾选第一行最左侧的复选框，点击“完成检索”，得到 250 篇引用文献（图 7-4）。

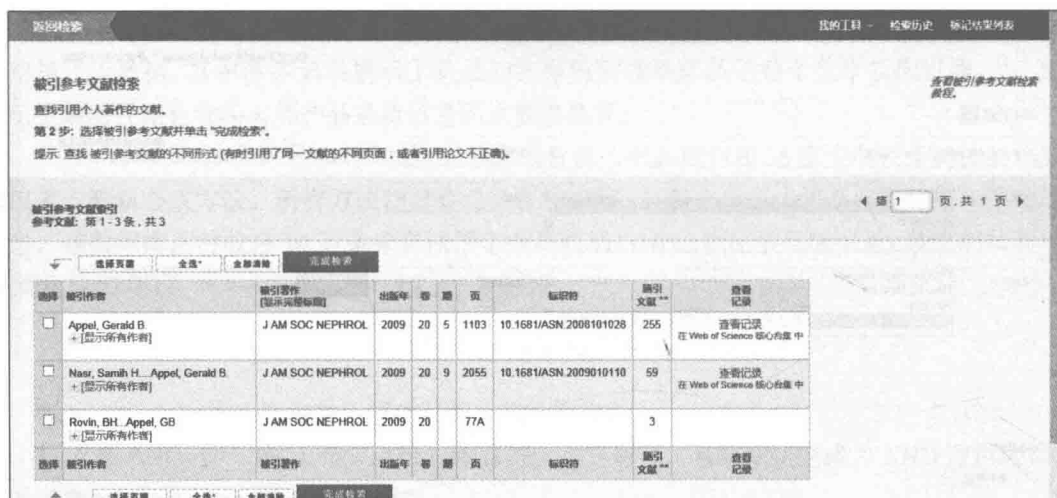


图 7-3 核心合集的被引文献索引

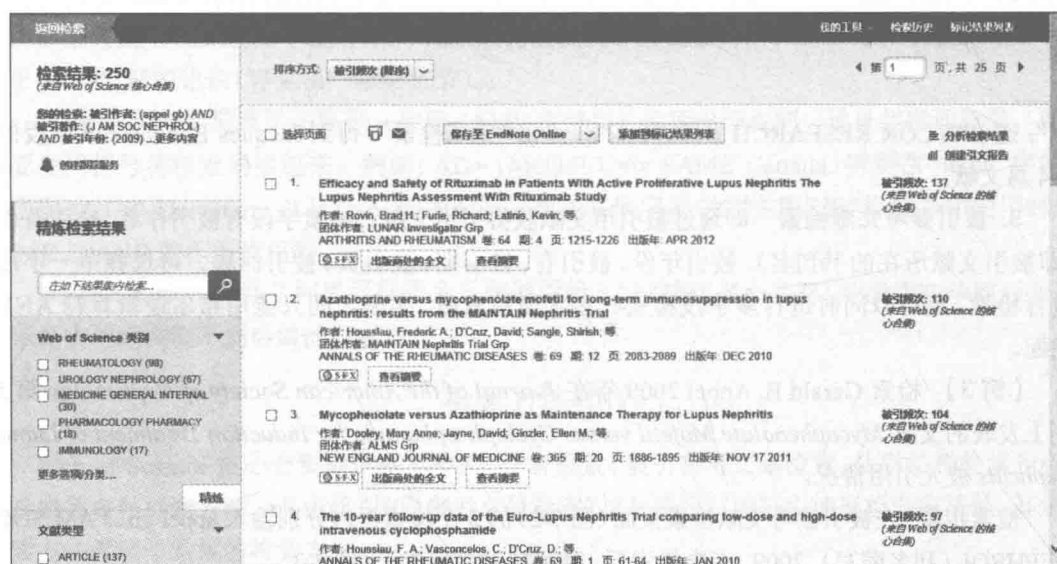


图 7-4 核心合集的引文检索结果

在图 7-3 页面上, 点击索引条目右侧的“查看记录”可见被引用文献的记录(图 7-5), 再通过右侧的“被引频次”前的数字, 也能得到引用文献一览。只有被 Web of Science 核心合集收录的文献, 才有“查看记录”链接。若想持续关注某一文献被引用情况, 可在来源文献记录中点击“创建引文跟踪”。

在对某一特定作者进行引文检索时, 要注意可能会出现同名同姓不同人的情况。可通过文献主题、期刊名称和作者机构的不同加以鉴别。

4. 化学结构检索 在 Index Chemicus 和 Current Chemical Reactions 2 个数据库中可以通过化学结构图或文本词检索化合物信息和化学反应信息。化学结构检索界面(图 7-6)分三个检索区域, 分别是: 化学结构图(通过画化学结构图检索)、化合物数据(通过化合物名称、分子量等检索)和化学反应数据(通过化学反应关键词、化学反应条件等检索)。

5. 高级检索 只限于来源文献检索, 不能进行引文检索。使用高级检索有以下 3 个优点(图 7-7):

(1) 可在检索提问框内直接输入带有字段标识符的检索词或检索式, 如 TS=(hepatitis B virus), TS 表示主题字段, 包含标题、关键词和文摘三个字段。



EndNote
全文检索 查找全文 保存至 EndNote Online 添加新记录到列表

我的工具 检索历史 标记结果列表
返回列表 第 1 页, 共 1 页

Mycophenolate Mofetil versus Cyclophosphamide for Induction Treatment of Lupus Nephritis

作者: Appel, GB (Appel, Gerald B)^{1,2,3}; Contreras, G (Contreras, Gabriel)^{4,5}; Dooley, MA (Dooley, Mary Anne)^{6,7}; Ginzler, EM (Ginzler, Ellen M)^{8,9}; Isenberg, D (Isenberg, David)¹⁰; Jayne, D (Jayne, David)¹¹; Li, LS (Li, Lai-Shi)¹²; Mysler, E (Mysler, Eduardo)^{13,14}; Sanchez-Guerrero, J (Sanchez-Guerrero, Jorge)^{15,16}; Solomon, N (Solomon, Neil)¹⁷ 更多内容
团体作者: Aspreva Lupus Management Study Grp

JOURNAL OF THE AMERICAN SOCIETY OF NEPHROLOGY
卷: 20 期: 5 页: 1103-1112
DOI: 10.1681/ASN.2008101026
出版年: MAY 2009
查看期刊信息

摘要

Recent studies have suggested that mycophenolate mofetil (MMF) may offer advantages over intravenous cyclophosphamide (IVC) for the treatment of lupus nephritis, but these therapies have not been compared in an international randomized, controlled trial. Here, we report the comparison of MMF and IVC as induction treatment for active lupus nephritis in a multinational, two-phase (induction and maintenance) study. We randomly assigned 370 patients with classes III through V lupus nephritis to open-label MMF (target dosage 3 g/d) or IVC (0.5 to 1.0 g/m² in monthly pulses) in a 24-wk induction study. Both groups received prednisone, tapered from a maximum starting dosage of 60 mg/d. The primary end point was a prespecified decrease in urine protein/creatinine ratio and stabilization or improvement in serum creatinine. Secondary end points included complete renal remission, systemic disease activity and damage, and safety. Overall, we did not detect a significantly different response rate between the two groups: 104 (56.2%) of 185 patients responded to MMF compared with 98 (53.0%) of 185 to IVC. Secondary end points were also similar between treatment groups. There were nine deaths in the MMF group and five in the IVC group. We did not detect significant differences between the MMF and IVC groups with regard to rates of adverse events, serious adverse events, or infections. Although most patients in both treatment groups experienced clinical improvement, the

引文网络

255 被引次数
35 引用的参考文献
查看 Related Records
查看引证关系图
创建引文图
(用 Web of Science 核心合集)

全部被引次数

271 / 所有数据集
255 / Web of Science 核心合集
144 / BIOSIS Citation Index
14 / 中国科学引文数据库
0 / Data Citation Index
3 / Sciendo Citation Index

最近的引文

Alexander, Suceana
Pharmacokinetics of Concentration-Controlled Mycophenolate Mofetil in Proliferative Lupus Nephritis: An Observational Cohort Study

图 7-5 核心合集收录的文献记录

化学结构检索

输入化学结构名称或任何化合物的数据。然后单击“检索”按钮进行检索。该检索结果将添加到检索历史中。查看化学结构检索历史。

检索 清除

化学结构的描述
通过化学名称或任何化合物的数据，在“检索”按钮下方输入“检索式”框中。然后选择检索模式。

子结构

结构匹配

检索式

C1CCN(C1)C2CCCCC2

化学物数据
输入化合物名称、生物活性或分子式。使用筛选条件进行筛选。

化合物名称:
化合物生物活性: 生物活性列表
分子式:
☐ 作为反应物 ☐ 作为产物 ☐ 作为催化剂 ☐ 作为试剂

化学反应数据
输入要检索的化学反应条件以及所需的关键信息。

气相色谱: ☐ 标记标记
其他: 详细列表
压力 (kPa):
时间 (小时):
温度 (摄氏度):
产率:
反应机理: 反应机理列表
化学键连接:

检索 清除

图 7-6 核心合集的化学结构检索界面

- (2) 可以保存检索历史，方便日后检索同一课题。
- (3) 可以对检索历史中先前用过的检索式进行组配检索。这一功能在分离文献他引和自引时更显优势。



图 7-7 核心合集的高级检索和检索历史

四、检索结果处理和分析

Web of Science 核心合集具有多种管理分析功能，这些管理分析功能在 WOS 其他数据库中多数也有，但与引文有关的功能只有在 Web of Science 核心合集中才有。

1. 输出记录 先勾选所需论文前面的复选框，再选择打印、E-mail 或保存至 EndNote Online 等输出方法。“保存至”的下拉菜单中有多种文献管理软件可供选择。如想保存至 Note Express 文献管理软件，选择“保存为其他文件格式”。

2. 保存历史 / 创建跟踪 通过高级检索中的保存历史 / 创建跟踪，可以保存检索历史以备日后调用，或者定期获得这一课题的最新文献。这项服务需免费注册后才能使用。操作步骤是：在检索历史页面上点击“保存历史 / 创建跟踪”→登录→完成表单的填写，最后点击“保存”。

检索历史可以选择保存在远地服务器上，也可以选择保存在本地的计算机上。定题跟踪服务创建成功后，所设定的电子邮箱每周或每月会收到与保存的检索式相匹配的新文献。需要注意的是，定题跟踪服务只限于与检索历史中最新一个检索式匹配的文献（检索式序号最大的），而不是与检索历史中所有检索式匹配的文献。



保存检索历史

检索历史名称: _____ (必填)

说明: _____ (可选)

电子邮件跟踪: ☒

电子邮件地址: _____

类型: 作者、标题、来源出版物

格式: 纯文本

频率: ☒ 每周 ☐ 每月

跟踪检索式: 主题=(avian influenza or bird flu)

☐ 保存检索历史后才可使用 RSS feed。

保存 取消

保存至本地磁盘

保存检索历史至本地磁盘。保存后, 关闭此窗口。

保存

图 7-8 核心合集的保存检索历史

如果只需要保存检索历史以供日后检索调用,而不需要创建定题跟踪服务,在图 7-8 中不要勾选“电子邮件跟踪”选项。

3. 创建引文跟踪服务 若对某篇文献被人引用的情况感兴趣,可创建引文跟踪服务。创建引文跟踪服务后,将来只要有人引用了所设定的文献,用户的电子邮箱就会接收到引用文献的信息。操作步骤是:在欲被跟踪文献的全记录显示页面,点击右侧的“创建引文跟踪”→登录→在弹出的新窗口中点击“创建引文跟踪”即可。

若要修改或删除已创建的引文跟踪服务,可以点击主页上的“我的工具”下拉菜单中的保存的检索式和跟踪,在新窗口可以修改和删除。

4. 分析/提炼检索结果 对检索结果进行分析有助于从宏观上把握检出文献的各种分布情况,通过“分析检索结果”或者“提炼检索结果”,可以回答以下问题:

(1) 某一专题的核心研究人员有哪几位(按作者排序)? 某一专题文献的高产国家有哪些(按国家/地区排序)? 某一专题的权威研究机构有哪些(按机构扩展排序)? 从而了解哪些学者、机构和地区在进行这个研究领域的工作。

(2) 某一专题的研究起始于什么年份或历史上研究的高峰期处什么年代(按出版年排序)? 了解文献被引用的高峰期是否已过。

(3) 某一专题文献主要集中在什么刊物上(按来源出版物名称排序)? 了解所关心的文献主要被哪些刊物引用? 以便确定日后期刊选读和投稿方向。

(4) 某一方面文献所属学科的分布情况和学科交叉情况如何(按 Web of Science 类别或研究方向排序)? 从中可了解某课题研究的学科交叉趋势。

分析检索结果比提炼检索结果功能更强大些,例如会显示每个选项在整个结果中所占的比例。

5. 创建引文报告 在检索结果列表页面上,点击右上角的“创建引文报告”,可对检索结果创建引文报告。图 7-9 是对“循环肿瘤细胞诊断结直肠癌转移”课题创建的引文报告(检索式为: ts = "circulating tumor cell*" AND ts = "colorectal cancer")。引文报告中的两个直方图显示检出文献近 20 年来每年发表的文献量和每年的引文量,显示课题的研究历程和被关注程度。

直方图右侧可见被引频次总计、去除自引的被引频次总计、施引文献、去除自引的施引文献、平均每篇被引次数和检出文献的 H 指数。

直方图下列出每篇检出文献的被引总次数,并由高到低排序。由此可获得这个课题的高被引文献。

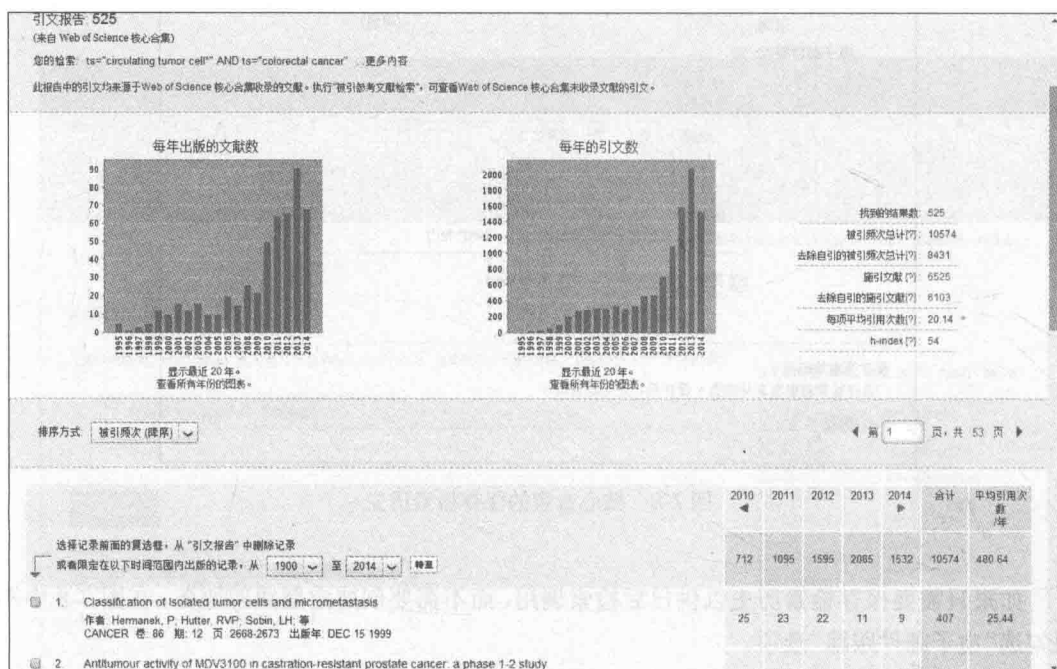


图 7-9 核心合集的引文报告

五、个性化功能和服务

用户免费注册并登录后,进入“我的工具”可使用系统提供的个性化服务。主要有:保存检索式和跟踪、EndNote 和 ResearcherID。

1. 保存检索式和跟踪 显示用户之前所创建的引文跟踪和保存的检索式,并且可以对感兴趣的期刊进行跟踪。定期获得相关的文献、期刊目录和检索结果(图 7-10)。

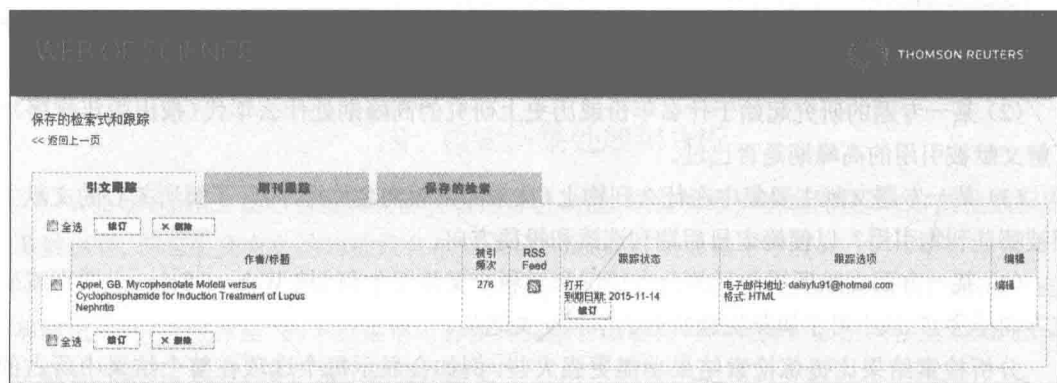


图 7-10 保存的检索式和跟踪

2. EndNote 是一种免费的网络版个人文献管理工具(图 7-11)。供存放用户检索结果,以便随时访问和使用。可创建符合各种出版样式的参考文献列表,方便撰写论文。

3. ResearcherID 提供全球每位学术研究人员一个专属的身份识别号码,让研究人员建立自己的著作清单,产生个人的引用信息,让全世界的学者看到自己的研究成果。同时,通过 ResearcherID 可以找到全世界的学者,了解他们的研究动态,与他们联系共同合作。

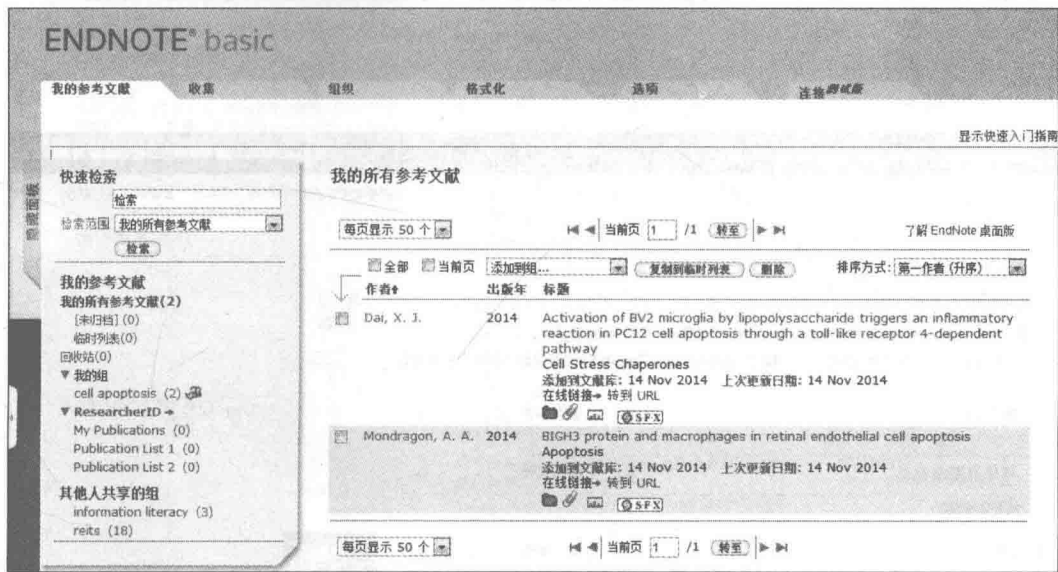


图 7-11 EndNote basic 主界面

第三节 中文引文检索

中文引文数据库有：中国科学引文数据库、中国引文数据库、中国生物医学期刊引文数据库、中文社会科学引文索引等。含有引文检索的其他中文数据库还有：中国生物医学文献数据库(SinoMed)、中国学术期刊网络出版总库、中文科技期刊数据库等。

一、中国科学引文数据库

1. 概况 中国科学引文数据库(Chinese Science Citation Database, CSCD)是由中国科学院文献情报中心创建, 2013—2014 年度收录我国数学、物理、化学、天文学、地学、生物学、农林科学、医药卫生、工程技术、环境科学和管理科学等领域出版的来源期刊 1141 种, 其中中国出版的英文期刊 125 种, 中文期刊 1016 种。中国科学引文数据库来源期刊分为核心库和扩展库两部分, 其中核心库 780 种, 累计从 1989 年到 2014 年 8 月底的论文记录近四百万条, 引文记录四千万条。

CSCD 是我国自行开发的第一个引文数据库。1995 年出版印刷版, 1998 年出版了光盘版, 2003 年推出网络版, 2007 年成为当时 ISI Web of Knowledge 平台(现在的 Web of Science 平台)上第一个非英文语种的数据库。

2. 检索方法 由于同在一个检索平台, CSCD 的检索方法与 Web of Science 核心合集类似。被引参考文献检索, 引文检索中的字段选项有: 被引作者、被引第一作者、被引来源、被引机构、被引实验室、被引文献主编。

【例 4】查找闻玉梅 2003 年发表在中华肝病杂志上的一篇文章的被引用情况。

在被引参考文献检索页面(图 7-12)的 3 个检索框内分别输入闻玉梅、中华肝病杂志、2003, 点击检索后, 得到被引参考文献索引(图 7-13)。

在被引参考文献索引中勾选第 2 条参考文献前面的复选框, 点击“完成检索”。得到引文检索的最终结果(图 7-14): 4 篇引用文献。

检索 中国科学引文数据库 S3 我的工具 检索历史 标记结果列表

欢迎使用全新的 Web of Science! 查看快速入门教程。

被引参考文献检索

查找引用个人著作的文献。

第 1 步: 输入有关被引著作的信息。各字段用布尔逻辑运算符 AND 相组配。

*注意: 输入与其他字段相组配的卷、期或页可能会降低检索到的被引参考文献不同形式的数量。

查看被引参考文献检索教程。

被引作者
 被引著作
 被引年份

查看编写列表

检索

+ 添加另一字段 | 清除所有字段

可以进行英文或简体中文检索 (正在进行中文检索)

图 7-12 CSCD 的被引参考文献检索

返回检索 我的工具 检索历史 标记结果列表

被引参考文献检索

查找引用个人著作的文献。

第 2 步: 选择被引参考文献并单击“完成检索”。

提示: 查找被引参考文献的不同形式 (有时引用了同一文献的不同页面, 或者引用论文不正确)。

查看被引参考文献检索教程。

被引参考文献索引
参考文献: 第 1-2 条, 共 2

第 1 页, 共 1 页

选择页面 全选* 全部清除 完成检索

选择	被引作者	被引著作 [显示完整标题]	出版年	卷	期	页	标识符	被引文献**	查看记录
☐	闻玉梅	中华肝脏病杂志	2003	9		519-521		1	
☐	Wen Yumei 闻玉梅	Chinese Journal of Hepatology 中华肝脏病杂志	2003	11	9	519	1007-3418(2003) 11:9<519-521:JYXG>2.0 TX;2-C	4	查看记录 在中国科学 引文数据库 中

选择 被引作者 被引著作 出版年 卷 期 页 标识符 被引文献** 查看记录

选择页面 全选* 全部清除 完成检索

图 7-13 CSCD 的被引参考文献索引





图 7-14 CSCD 的引文检索结果

二、中国引文数据库

中国引文数据库 (Chinese Citation Database, <http://ref.cnki.net>) 是中国知网的一个子数据库。引文数据来源于中国学术期刊网络出版总库、中国博士学位论文全文数据库、中国优秀硕士学位论文全文数据库、中国重要会议论文全文数据库、中国重要报纸全文数据库、中国图书全文数据库和中国年鉴全文数据库等。

1. 中国引文数据库的优点 该数据库的引文检索功能比国内其他引文数据库完善。国内许多图书馆在开展查收查引服务中, 将该数据库作为中文引文检索的必查数据库。其优点如下: ①引文检索的检索项细化为被引题名、被引作者、被引第一作者、被引关键词、被引摘要、被引单位、被引刊名、被引年、被引期、被引基金、被引 ISSN、被引统一刊号共 12 个选项, 可满足不同角度的引文检索要求; ②由于收录范围广, 检索得到的引用文献除期刊论文外, 还有学位论文、会议文献等。

2. 检索方法 中国引文数据库的检索方法主要有快速检索和高级检索。

(1) 快速检索: 为该数据库的默认检索页面, 检索提问框内可输入被引作者、被引刊名书名、被引题名词等, 检索返回页面显示被引用文献题录, 继续点击被引文献篇名, 得到被引文献的摘要, 点击“本文的引文网络”中的“引证文献”, 即可获得引用文献的题录, 点击篇名即可获得全文。

(2) 高级检索: 分“源文献检索”和“引文检索”。“引文检索”提供多个检索提问框和有 12 个被引字段选择的下拉菜单, 检索时间限制也细化为文献发布时间和文献被引时间。高级检索中引文检索的步骤是: 检索到被引文献→通过被引频次的数字链接得到引用文献题录。若要看引用文献的全文, 按被引文献题名→引证文献题名→全文链接这 3 个步骤进行。

【例 5】通过高级检索查找吴孟超 1991 年发表于《实用外科杂志》上《原发性肝癌的治疗原则》一文的被引情况。

检索步骤是: <http://ref.cnki.net> → 点击“高级检索” → 输入检索词, 点击“检索” → 点击图 7-15

中的“原发性肝癌的治疗原则”→得到被引文献的文摘格式(图 7-16), 点击“本文的引文网络”中的“引证文献(31)”, 即可获得 31 篇引用文献的题录, 点击篇名即可获取全文。



图 7-15 中国引文数据库高级检索中的引文检索



图 7-16 被引文献的文摘格式和引文网络

本例检索到的 31 篇引证文献(图 7-17), 28 篇来自于中国期刊全文数据库, 1 篇来自于中国博士学位论文全文数据库, 2 篇来自于中国优秀硕士学位论文全文数据库。只有同时是相应全文数据库的授权用户, 才能看到来源于不同子数据库的全文。



图 7-17 引用文献的题录格式

通过高级检索中的“源文献检索”, 也能进行引文检索。检索步骤是: 在“源文献检索”中先查到某篇文献→点击文献标题→点击“本文的引文网络”中的“引证文献”, 即可获得引用文献的题录。用“源文献检索”进行引文检索的缺点是: 未曾被引用过的文献也会出现在检索结果中。如果本文的引文网络中的“引证文献”数字为 0, 即可认定该文在中国知网统计范围内尚未被引用。

三、其他中文数据库的引文检索

1. 中国生物医学文献数据库 是中国医学科学院医学信息研究所开发研制的综合性医学文献数据库。SinoMed 提供引文检索途径。检索字段有被引文献题名、被引文献主题、被引文献作者、被引文献第一作者、被引文献出处、被引文献机构、被引文献第一机构和被引基金等。

【例 6】检索林志雄 1998 年发表在《中华骨科杂志》上的论文《关节镜诊断和治疗膝关节骨关节炎》一共被引用了多少次。

检索时先点击“引文检索”, 第一个检索字段选“被引文献作者”, 输入林志雄, 第二个检索字段选被引文献题名, 输入“关节镜诊断”, 第三个字段选择被引文献出处, 输入“中华骨科杂志”, 发表年代选 1998—1998, 最后点击“检索”(图 7-18)。

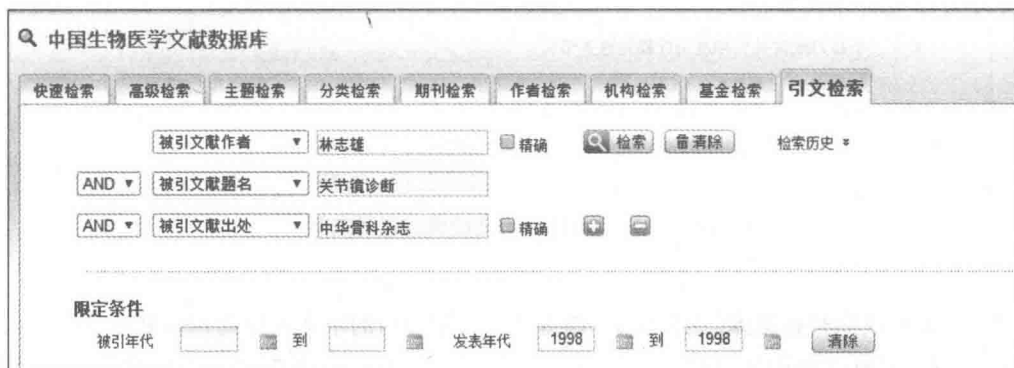


图 7-18 SinoMed 的引文检索界面

检索结果为林志雄的这篇被引用文献(图 7-19)。点击被引频次后面的数字 78, 即可获得这 78 篇引用文献(图 7-20)。

快速检索 高级检索 主题检索 分类检索 期刊检索 作者检索 机构检索 基金检索 引文检索

常用字段 [] 二次检索 [] 检索 [] 清除

检索条件: (((林志雄[被引文献作者]) AND (关节镜诊断[被引文献题名])) AND (中华骨科杂志[被引文献出处]))

限定条件: 1998-1998[发表年代]

被引频次大于 [请选择] 次 被引年代 [] 到 [] 发表年代 1998 到 1998

全部: 1

每页 20条 排序 被引频次 [] 引文追踪 结果输出 []

当前页 首页 | 上一页 | 下一页 | 尾页 共 1 篇 1 / 1 转页

1. 关节镜诊断和治疗膝关节骨关节炎

原文索取 [] 我的数据库 []

作者: 林志雄; 余楠生; 卢伟杰

出处: 中华骨科杂志 1998; 18(4): 199-202

被引频次: 78 参考文献数: 2

图 7-19 SinoMed 引文检索的被引用文献

中国生物医学文献数据库

快速检索 高级检索 主题检索 分类检索 期刊检索 作者检索 机构检索 基金检索 引文检索

关节镜诊断和治疗膝关节骨关节炎 的施引文献:

- 关节镜诊治膝关节骨关节炎临床探讨

作者: 李军; 马广文; 周盛智; 吕欧; 张杰

出处: 安徽医学 2010; 31(8): 924-926

参考文献数: 6
- 经关节镜膝关节手术的护理配合

作者: 秦晓梅

出处: 中华全科医学 2010; (10): 1335-1336

参考文献数: 9
- “软系”疗法在膝关节炎中的应用

作者: 丛明; 胡玉龙; 宋金华

出处: 中国疗养医学 2010; 19(9): 628-629

参考文献数: 4
- 小针刀配合关节冲洗治疗膝骨性关节炎

作者: 张英杰

出处: 国际中医中药杂志 2009; 31(2): 148-149

参考文献数: 2

图 7-20 SinoMed 引文检索的引用文献

2. 中文科技期刊数据库(引文版) 维普资讯有限公司的中文科技期刊数据库(引文版)提供引文检索, 又称文献引证追踪。这种检索方式先输入检索词(图 7-21), 查到被引文献, 然后点击该被引文献的被引量(图 7-22), 即可获得引用文献(图 7-23)。



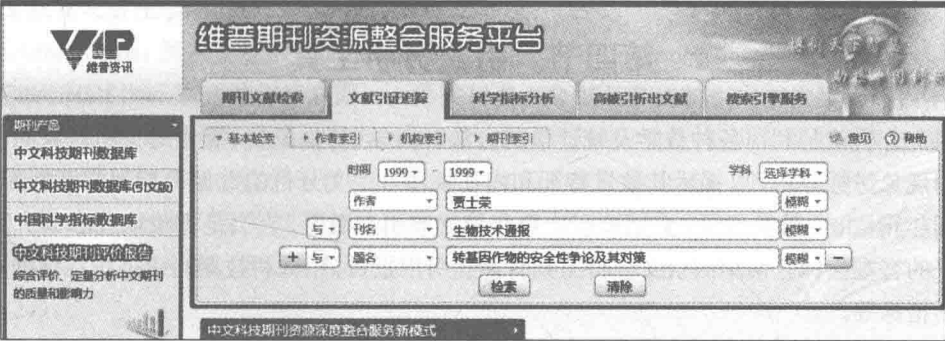


图 7-21 文献引证追踪的检索界面

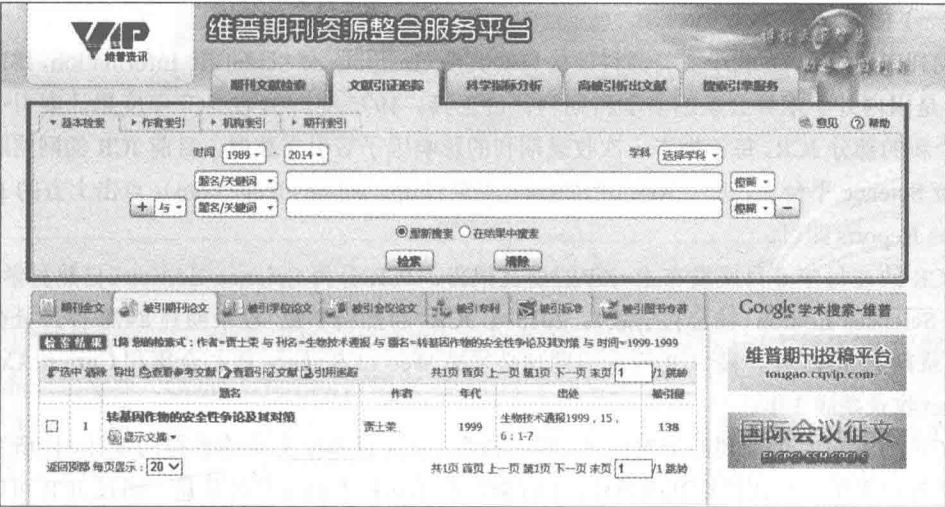


图 7-22 文献引证追踪被引文献



图 7-23 文献引证追踪引用文献

第四节 引文分析工具

引文分析就是利用各种数学及统计学的方法对科学期刊、论文、著者等分析对象的引用和被引用现象进行分析,以揭示其数量特征和内在规律。引文分析的数据有期刊的影响因子、期刊的即年指数、文献发表量、文献被引次数和篇均被引次数等,它们是评价期刊和研究人员学术水平的客观指标。常用的引文分析工具有期刊引用报告、中国科技期刊引证指标数据库和基本科学指标等。

一、期刊引用报告

(一) 概述

期刊引用报告(JCR),由美国科技信息研究所(Institute of Scientific Information, ISI)编辑出版,是国内外学术界公认的多学科期刊评价工具。1975年,ISI在SCI年度累计索引中增加了一个新的部分JCR,每年报告一次收录期刊的影响因子等引文数据。目前JCR的网络版集于Web of Science平台上,进入Web of Science主页(<http://isiknowledge.com>),点击上方的Journal Citation Reports即可。

JCR约在每年6月底发布上一年的引文报告。JCR分为Science Edition(自然科学版)和Social Sciences Edition(社会科学版)。2013年JCR“自然科学版”收录期刊8478种,“社会科学版”收录期刊3045种。被JCR收录的期刊必须被Web of Science核心合集和Current Contents Connect收录期满3年。

JCR通过对收录期刊的影响因子、即年指数、引文量和发文量等数据的统计,供研究人员和科研管理人员分析比较期刊的质量,了解期刊在本学科中的排名等信息。通过JCR可以查询到:哪些期刊被引用的频率最高?哪些期刊在本学科中影响力最大?哪些期刊发表文章最多?哪些学科文献被利用最广泛等。

(二) 期刊数据

通过JCR中的View Journal Data,可以了解一种期刊被引用的情况和在本学科中的影响力以及老化速度等。

1. JCR提供的期刊数据

Abbreviated Journal Title: 期刊刊名的缩写。

ISSN: 国际标准连续出版物号。

Total Cites: 被引总次数——某刊在JCR统计年被引文献总次数。

Impact Factor: 期刊影响因子——是指某刊前两年发表的论文在评价当年平均被引用的次数。即把某刊前两年发表的文献在当年被引用的次数作为分子,该刊前两年发表的文献数作为分母得出的数值。例如期刊*Nature Immunology* 2013年影响因子的统计方法是:该刊2011年和2012年发表的文献在2013年被引用次数分别为3913次和2580次,合计为6493次,该刊2011年和2012年在统计范围内的文献发表数分别是132篇和128篇,合计为260篇,前者除以后者等于24.973,即该刊2013年的影响因子为24.973。

影响因子是评价学术期刊质量最重要的指标。影响因子越高,表示该期刊文献的平均利用率越高,由此可推理出期刊的质量越高。利用期刊的影响因子,可以帮助选购馆藏期刊,确定各学科的核心期刊,指引期刊选读和投稿。

5-Year Impact Factor: 5年影响因子——期刊前5年发表的文献在评价当年被平均引用的次数。如果5年IF小于IF(2年期),表明期刊受关注度增加,反之则表示减少。

Immediacy Index: 即年指数、即时指数——期刊当年发表的文献在当年平均被引用的次数,



提示文献发表后在学术界所引起的反应速度。

Citable Items: 发表文献数——不包括 editorials, letters, meeting abstracts 等。

Cited Half-life: 期刊被引半衰期——将某刊在 JCR 统计年内被引用的次数依出版年份降序排列, 前 50% 论文的出版时间段即为该期刊的被引半衰期。期刊被引半衰期是测定某期刊文献老化速度的重要指标。被引半衰期越短, 期刊文献的老化速度越快, 反之则越慢。

2. 期刊数据检索 如果了解某种期刊的数据, 首先选“JCR Science Edition”, 并在右侧的下拉菜单中选择年份, 然后选右面的三个检索入口 (Select an option) 之一, 点击“SUBMIT” (图 7-24)。

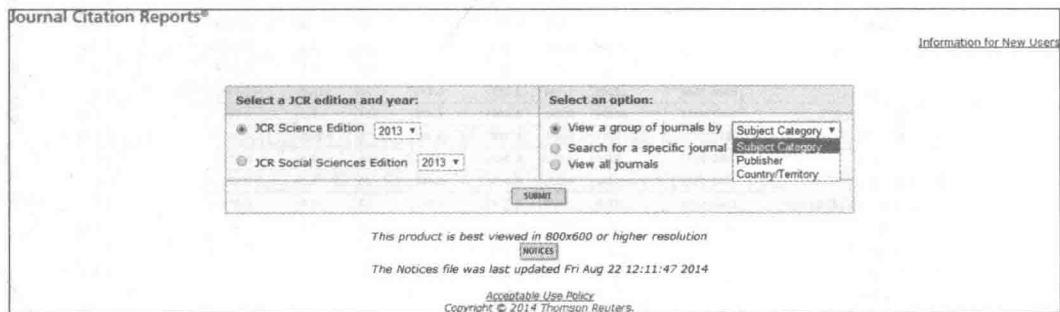


图 7-24 JCR 检索主页面

(1) View a group of journals by: 该检索入口又分有 Subject Category (主题分类, 又称学科)、Publisher (出版商)、Country/Territory (国家/地区) 3 个选项。

【例 7】通过 View Journal Data 查找 2013 年生物学领域影响因子最高的期刊是哪种。

检索步骤如下: 在图 7-24 中选“JCR Science Edition”, 年份选“2013”→检索途径选 View a group of journals by 中的 Subject Category, 点击 SUBMIT →在图 7-25 中的 Select one or more categories from the list 中选 BIOLOGY, 在 View Journal Data 的 sort by 下拉菜单中选 Impact Factor, 点击“SUBMIT”→得到 83 种生物学方面的期刊。2013 年影响因子最高的生物学期刊是 *PLoS Biology*, 影响因子为 11.771 (图 7-26)。

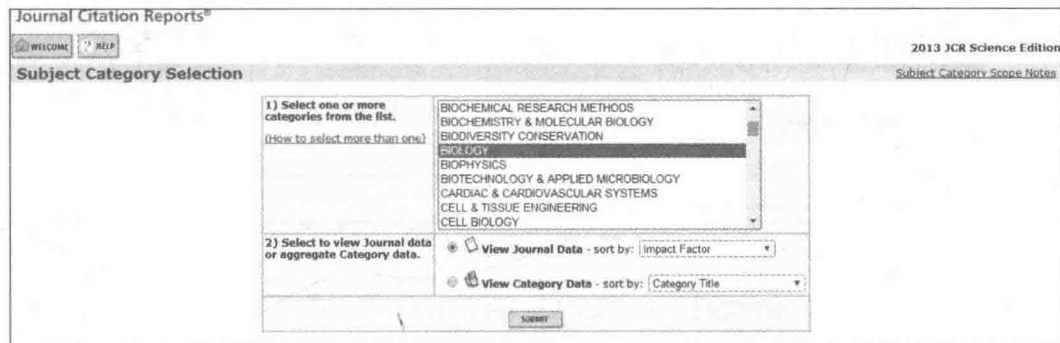


图 7-25 JCR 中的主题分类检索

继续点击图 7-26 上的期刊名 PLOS BIOL (刊名缩写), 可得到期刊的详细信息 (图 7-27)。

在详细信息的基本数据表中列出刊名全称、缩写、ISSN、被引总次数、影响因子、5 年影响因子、即年指数、发表文献数等各类数据, 下方有上述各类数据的详细计算过程。

另外, 在基本数据表的下方有被引期刊数据 (CITED JOURNAL DATA)——哪些期刊引用本刊的文章; 引用期刊数据 (CITING JOURNAL DATA)——本刊正在引用哪些期刊上的文章; 影响因子趋势图 (IMPACT FACTOR TREND)——展示本刊过去 5 年的影响因子; 相关期刊 (RELATED JOURNALS)——哪些期刊由于曾经彼此引用而与本刊相关等。

Journal Citation Reports®

2013 JCR Science Edition

Journal Title Changes

Journal Summary List

Journals from: subject categories BIOLOGY

Sorted by: Impact Factor

Journals 1 - 20 (of 83)

Ranking is based on your journal and sort selections.

Mark	Rank	Abbreviated Journal Title (linked to journal information)	ISSN	JCR Data (J)					Eigenfactor® Metrics (J)		
				Total Cites	Impact Factor	5-Year Impact Factor	Immediacy Index	Articles	Cited Half-life	Eigenfactor® Score	Article Influence® Score
☐	1	PLOS BIOL	1545-7885	24324	11.771	12.807	1.706	201	5.9	0.11227	7.042
☐	2	BIOL REV	1464-7931	7261	9.790	10.912	2.071	56	8.7	0.01426	4.179
☐	3	PHYS LIFE REV	1571-0645	828	9.478	8.322	12.167	12	3.5	0.00274	2.778
☐	4	ELIFE	2050-084X	722	8.519	8.556	2.127	228	0.7	0.00277	6.134
☐	5	BMC BIOL	1741-7007	3214	7.431	6.942	1.393	84	4.0	0.01637	2.853
☐	6	PHILOS T R SOC B	0962-0436	29669	6.314	7.960	3.026	307	7.1	0.08509	3.259
☐	7	FASER J	0892-6638	41104	5.480	6.045	1.371	466	8.5	0.07541	2.092
☐	8	BIOSCIENCE	0006-3568	12300	5.439	6.727	0.543	70	>10.0	0.01852	2.961
☐	9	P ROY SOC B-BIOL SCI	0962-0452	39399	5.292	5.808	1.025	513	8.2	0.09359	2.318
☐	10	Q REV BIOL	0033-5770	3503	5.059	7.358	0.333	9	>10.0	0.00273	3.079
☐	11	BIOESSAYS	0265-9247	8726	4.838	4.747	1.033	122	8.6	0.01958	1.973
☐	12	BIOL DIRECT	1745-6150	1274	4.035	3.463	0.500	30	4.0	0.00740	1.574
☐	13	BIOELECTROCHEMISTRY	1567-5394	3719	3.870	3.774	0.827	52	6.6	0.00595	0.822
☐	14	GEOBIOLOGY	1472-4677	1434	3.692	4.146	1.100	40	4.9	0.00511	1.542
☐	15	BIOL LETTERS	1744-9561	6445	3.425	3.902	0.586	266	4.4	0.02874	1.560
☐	16	J BIOL RHYTHM	0748-7304	2571	3.316	3.500	0.605	38	9.2	0.00542	1.373
☐	17	INTERFACE FOCUS	2042-8898	496	3.124	3.153	0.939	66	2.0	0.00301	1.315
☐	18	J EXP BIOL	0022-0940	26378	3.007	3.235	1.034	508	0.4	0.04054	1.020

图 7-26 JCR 中的期刊概要列表

Journal Citation Reports®

2013 JCR Science Edition

Journal: PLOS BIOLOGY

Mark	Journal Title	ISSN	Total Cites	Impact Factor	5-Year Impact Factor	Immediacy Index	Citable Items	Cited Half-life	Citing Half-life
☐	PLOS BIOL	1545-7885	24324	11.771	12.807	1.706	201	5.9	7.0

Cited Journal (00) Citing Journal (00) Source Data Journal Self Cites

CITED JOURNAL DATA CITING JOURNAL DATA IMPACT FACTOR TREND RELATED JOURNALS

Journal Information (J)

Full Journal Title: PLOS BIOLOGY
ISO Abbrev. Title: PLoS. Biol.
JCR Abbrev. Title: PLOS BIOL
ISSN: 1545-7885
Issues/Year: 12
Language: ENGLISH
Journal Country/Territory: UNITED STATES
Publisher: PUBLIC LIBRARY SCIENCE
Publisher Address: 1160 BATTERY STREET, STE 100, SAN FRANCISCO, CA 94111
Subject Categories: BIOCHEMISTRY & MOLECULAR BIOLOGY

Eigenfactor® Metrics
Eigenfactor® Score
0.11227
Article Influence® Score
7.042

Additional Links
GO TO CC CONNECT
CSDL Union

Journal Rank In Categories: (J) JOURNAL RANKING

Journal Impact Factor (J)

Cites in 2013 to items published in: 2012 = 1342 Number of items published in: 2012 = 152
2011 = 2566 2011 = 180
Sum: 3908 Sum: 332

Calculation: Cites to recent items 3908 = 11.771
Number of recent items 332

5-Year Journal Impact Factor (J)

图 7-27 JCR 中期刊的详细信息

(2) Search for a specific journal: 该途径用于检索已知的特定期刊, 检索词可以是刊名全称、刊名缩写、刊名中的单词或期刊的 ISSN。

(3) View all journals: 浏览查询全部期刊。

(三) 学科数据

通过 JCR 中的 View Category Data, 可以了解一个学科文献被利用程度和学科的知识老化速度等, 还可对学科之间文献被利用程度等进行比较。JCR 的学科数据有:

Total Cites (总引用次数)——某学科中期刊被引文献总次数。



Median Impact Factor(中值影响因子)——“中值影响因子”取自于一个学科中影响因子排序居中的那个期刊的影响因子。若某一学科的期刊为双数时,取影响因子居中的两种期刊影响因子的平均数为该学科的中值影响因子。

Aggregate Impact Factor(学科集合影响因子)——指前两年本学科期刊上的文献在 JCR 统计当年被平均引用的次数,即分子是该学科期刊前两年发表的文献在统计当年被引用的次数,分母是该学科前两年发表文献的总数。

Aggregate Immediacy Index(学科集合即年指数)。

Aggregate Half-Life(学科集合被引半衰期)。

Journals——该学科被 JCR 收录的期刊数。

Articles——该学科被 JCR 收录期刊上发表的文献数。

(四) 检索结果处理

JCR 检索结果的输出有打印和保存两种方式,操作步骤是:在检出的期刊概要列表(图 7-26)中,在需要输出期刊的“Mark”复选框内打勾,点击“Update Marked List”→点击“Marked List”→点击“Save to file”进行保存,或通过“Format for Print”进行打印。

二、基本科学指标

1. 概述 基本科学指标(ESI)是 ISI 于 2001 年推出的衡量科研绩效的分析评价工具。ESI 中的统计数据来源于 SCIE 和 SSCI 收录的 10 000 多种学术期刊,每两个月数据更新一次。ESI 主要从引文分析角度对国家、机构、期刊、论文、科学家进行统计和排序,统计指标包括文献发表量、文献被引次数、篇均被引次数。通过这些引文数据,可看各学科高被引论文(highly cited paper)和热门论文(hot paper),还可了解特定论文在本学科中的被引百分位档次(详见第十一章第二节文献信息分析)。ESI 默认的排序指标是文献被引次数。

ESI 分 22 个学科,它们是农业科学、生物与生化、化学、临床医学、计算机科学、经济学与商务、工程、环境与生态学、地球科学、免疫学、材料科学、数学、微生物学、分子生物学与遗传学、综合交叉学科、神经科学与行为学、药理学与毒理学、物理学、植物与动物科学、精神病学与心理学、社会科学、太空科学。

近年来,ESI 的科研评价作用已引起国内学界和科研管理部门的重视,不乏有人利用其丰富的数据立题研究,进行各机构科研实力对比和文献计量分析。

2. ESI 的检索途径 ESI 的检索主界面上提供 3 种检索途径(图 7-28)。

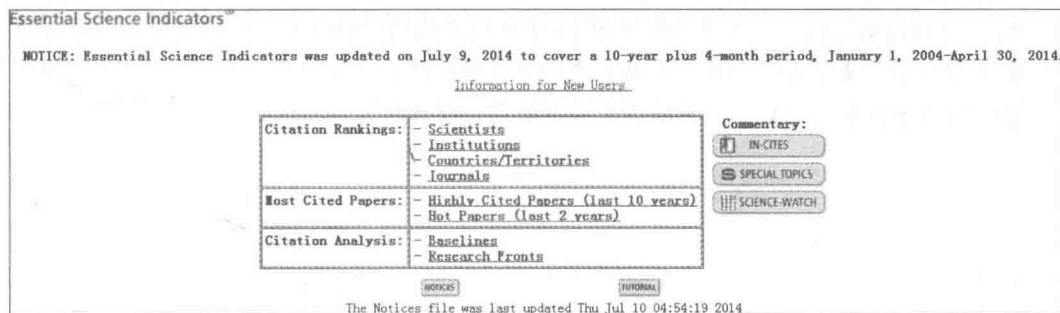


图 7-28 ESI 主界面

(1) Citation Rankings(引文排名):从 4 个角度提供引文排名,Scientists(取排名前 1%)、Institutions(取排名前 1%)、Countries/Territories(取排名前 50%)、Journals(取排名前 50%)。

(2) Most Cited Papers:分有 Highly Cited Papers(高被引论文)和 Hot Papers(热门论文),前者是近十年发表各领域被引排名前 1% 的文章,后者是近两年发表各领域在最近两个月内

被引排名前 0.1% 的文章。

(3) Citation Analysis (引文分析): 包括 Baselines 和 Research Fronts Baselines 下分三块: ① View the average citation rates table——10 年间各学科中每年发表论文的篇均被引次数; ② View the percentiles table——给出某一学科近 11 年中某一年的文献进入前 0.01%、0.1%、1%、10%、20%、50% 所必须达到的引文量。③ View field rankings table——给出 22 个学科被引量排名。Research Fronts 按照共引关系聚类高被引论文, 以揭示各学科的研究前沿, 帮助发现隐含的突破性研究。

在 ESI 主界面 (图 7-28) 右侧的 Commentary 下有 3 个选项: ① IN CITES——可见对各学科高被引科学家和机构等的采访、介绍和评论文章; ② SPECIAL TOPICS——介绍一部分选定研究领域的引文分析和统计的方法; ③ SCIENCE-WATCH——通过采访、评论、方法介绍等反映 ESI 统计报告背后的知识和信息。

本章小结

本章简要介绍了引文检索的相关概念和作用, 重点通过实例对 Web of Science 核心合集、中国科学引文数据库、中国引文数据库等常用引文数据库的检索方法、检索结果分析和创建引文跟踪服务等管理分析功能作了详细的说明。最后介绍了引文分析工具——期刊引用报告和基本科学指标的使用。

(符礼平)

思考题

1. 引文检索有哪些作用?
2. 对于一篇没有参考文献的文章, 能查到该文被引用情况吗?
3. 在 Web of Science 平台上, 有引文检索功能的数据库有哪几个?
4. Web of Science 核心合集由哪几个子数据库构成?
5. Web of Science 核心合集有哪几条检索途径?
6. 一篇文献所载期刊未被 Web of Science 核心合集收录, 通过其基本检索途径, 能查到这篇文献吗? 通过其被引参考文献检索有可能查到该篇文献被引用情况吗?
7. 创建定题跟踪服务和创建引文跟踪服务有什么异同?
8. 与其他有引文检索的中文数据库相比, 中国引文数据库有哪些优点?
9. JCR 中最重要的期刊数据是什么? 查询该数据有什么用处?
10. 2013 年 JCR 中免疫学的影响因子最高期刊是哪本?



第八章 专类信息检索

传统意义上的医学文献,就文献类型而言主要指医学图书和期刊,从学科的角度主要涵盖基础医学、临床医学和预防医学等。由此,常将之外的文献统一归入特种文献的范畴。随着科学的进步、时代的发展,尤其是信息环境的变化,对医学文献信息的划分亦在悄然演化。围绕某一新兴学科或边缘性领域的专类文献信息如循证医学和转化医学,针对某种特定目的如学位论文和药品说明书,基于知识产权保护理念的如专利文献,源于特定源头的如会议文献和会议信息,借助于特殊媒介的如视听文献等,因对其范畴的专门限定而将其归之为专类信息。对这类信息的检索,既遵循信息获取与利用的共有规律与原则,更有其各自独特的方式与路径。因此,本章将介绍学位论文、会议信息、专利文献、循证医学信息、药学(品)信息、疾病知识库以及视频数据库等各类中外常用的专类文献信息资源系统的特点及其检索方法。

第一节 学位论文检索

一、概 述

学位论文是学生为获得学位向高等院校或其他学术研究机构提交的学术研究论文,包括学士论文、硕士论文、博士论文等。目前数据库建设主要是针对硕士和博士学位论文。

学位论文具有科研论文的科学性、学术性、新颖性、绝大多数不公开发表或出版等特性。其中博硕士论文的特点为:①选题新颖,带有独创性;②问题专一,论述系统;③来源分散,大多不正式出版,使用难度较大;④具有连续性和继承性。

学位论文由如下部分组成:①封面;②版权声明;③题目;④中文摘要;⑤英文摘要;⑥目录;⑦序言(或序论、导论);⑧正文;⑨注释;⑩结论;⑪参考文献;⑫附录;⑬作者的致谢、后记或说明等一律列于论文末尾;⑭学位论文原创性声明和授权使用说明;⑮封底。

学位论文数据库主要有:①万方数据《中国学位论文全文数据库》;②CNKI《中国优秀博硕士学位论文全文数据库》;③CALIS高校学位论文库等。

二、国内学位论文检索

(一) 万方数据《中国学位论文全文数据库》

万方数据《中国学位论文全文数据库》由中国科技信息研究所提供。中国科技信息研究所是国家法定的学位论文收藏机构,各高等院校、研究生院及研究所均向该机构送交硕士、博士和博士后的论文,收录了自1980年以来的论文近150万篇。涵盖自然科学、数理化、天文、地球、生物、医药、卫生、工业技术、航空、环境、社会科学、人文地理等各学科领域(图8-1)。

其中“985”及“211”高校论文收录量占总量的70%以上,每年增加约30万篇。

(二) 中国知网《中国优秀博硕士学位论文全文数据库》

CNKI《中国优秀博硕士学位论文全文数据库》是目前国内相关资源最完备、高质量、连续

动态更新的学位论文全文数据库。收录 1984 年至今(部分收录 1999 年之前)全国 621 家硕士培养单位的优秀学位论文及全国 404 家博士培养单位的博士论文,近 170 万篇。产品分为十大专辑,十大专辑下分为 168 个专题(图 8-2)。

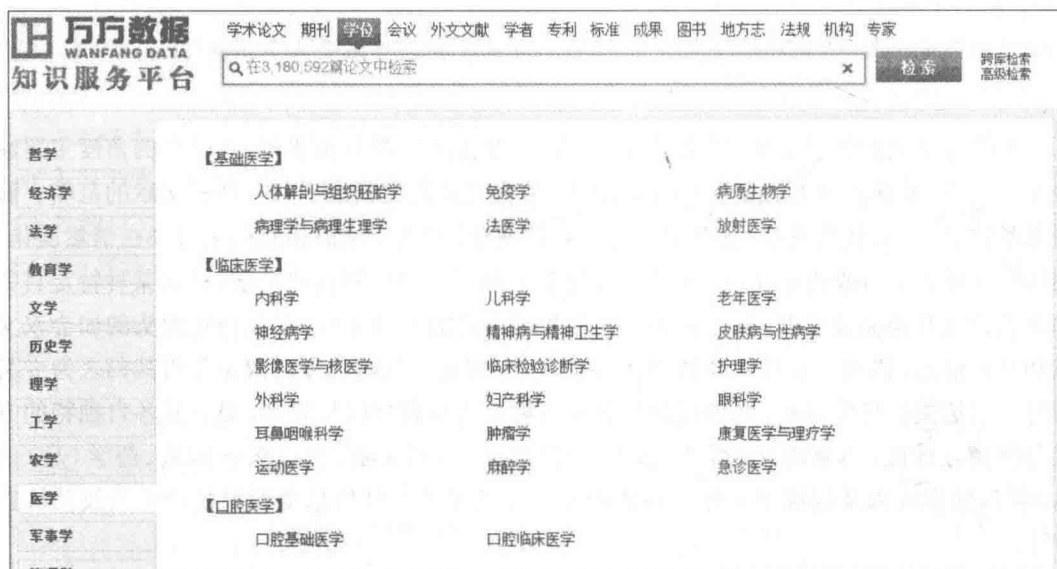


图 8-1 中国学位论文全文数据库检索界面

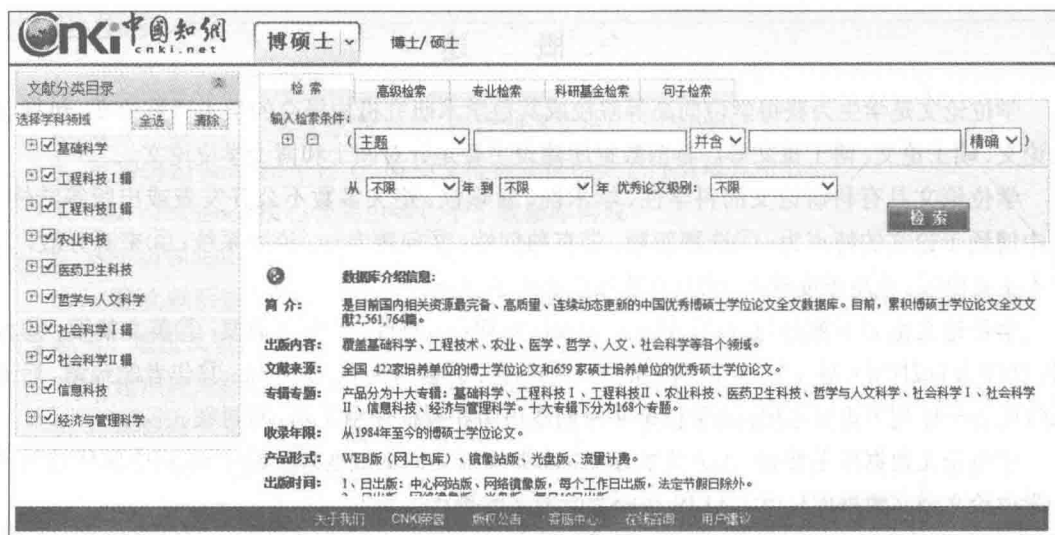


图 8-2 中国优秀博硕士学位论文全文数据库检索界面

(三) CALIS 高校学位论文库

CALIS 高校学位论文库 (<http://www.calis.edu.cn>) 收录包括北京大学、清华大学等全国著名大学在内的近百个 CALIS 成员馆的硕士、博士学位论文,收录加工数据已达 25.8 万条。该库只收录题录和文摘,没有全文,全文服务通过 CALIS 的馆际互借系统提供。内容涵盖自然科学、社会科学、医学等各个学科领域。用户可从题名、论文作者、导师、作者专业、作者单位、摘要、分类号、主题和全字段等不同角度进行检索(图 8-3)。



图 8-3 CALIS 学位论文库检索界面

三、国外学位论文检索

(一) ProQuest 国外学位论文检索系统

ProQuest 公司是世界上最早及最大的博硕士论文收藏和供应商,该公司的学位论文数据库(ProQuest Digital Dissertation, PQDD)收录的范围涉及美国、加拿大、欧洲及世界其他国家的 2000 多所大学和科研机构的文、理、工、农、医等领域。它是学术研究中十分重要的参考信息源,每年约增加 4.7 万篇博士论文和 1.2 万篇硕士论文摘要(图 8-4)。

国内部分高校图书馆、文献收藏单位联合购买了一定数量的 ProQuest 学位论文全文(PDF 格式),并提供网络共享,凡参加联合订购成员馆均可共享整个集团订购的全部学位论文资源。收录年限从 1980 年开始,每年增加 1 万多篇。

ProQuest 学位论文全文系统已在国内设立了 3 个镜像网站,可以同步访问。

CALIS 镜像站: <http://202.120.13.45/umi/index.jsp>。

中国科学技术信息研究所镜像站: <http://168.160.16.198/umi>。

(二) NDLTD 学位论文库

NDLTD (<http://www.ndltd.org>)是由美国国家自然科学基金支持的一个网上学位论文共建共享项目,为用户提供免费的学位论文文摘,还有部分可获取的免费学位论文全文(根据作者的要求,NDLTD 文摘数据库链接到的部分全文分为无限制下载,有限制下载,不能下载几种方式),以便加速研究生研究成果的利用(图 8-5)。

目前全球有 170 多家图书馆、7 个图书馆联盟、20 多个专业研究所加入了 NDLTD,其中 20 多所成员已提供学位论文文摘数据库 7 万条,可以链接到的论文全文大约有 3 万篇。

ProQuest 学位论文全文检索平台

Back to Homepage Site: 中值所
English

Welcome!

Your Location:

Related Links

ProQuest
Beijing Zhongke I/E
Company

Homepage >> Advanced Search

Advanced Search

Search criteria

Title Include ALL words AND

Abstract Include ALL words AND

Subjects Include Select AND

Author Include ALL words AND

School Include ALL words AND

Advisor Include ALL words AND

Source Include ALL words

Publish Year: To

Degree: ☒ All ☐ Doctor ☐ Master

Language: All

Display: ☒ All ☐ Subscribed Full Text only

Search

图 8-4 ProQuest 学位论文检索界面

NDLTD Driven by Chamo Discovery

Login Cart Heading Search Clear Session

Search Advanced Search

Enter search terms below.

Any Field has all of these words

Any Field has this exact phrase

Any Field has at least one of these words

Any Field has none of these words

You may also limit the results using the filters below.

Language

Afar
Afrikaans
Akkadian
Albanian
Amharic

Format

Adobe Acrobat PDF
Binary File
GIF Image
Gnu ZIP Archive
HTML Document

Version 1.131

图 8-5 NDLTD 学位论文库检索界面

(仇晓春)

第二节 会议文献检索

一、概 述

学术会议是一种非常重要的学术交流方式,专业人员可借此交流思想、传递信息、相互启迪与借鉴。会议论文和期刊论文在学术交流中起着互补的作用,会议论文是专家学者表述其初始思路,从同行中收集反馈,改善进一步的研究的报道。医学会议是医学科研人员交流与传递医学信息的重要途径。由于一些新问题、新见解、最新研究成果或进展等大多数都是在会议上首次被提出来的,及时地获取和掌握这些信息对于医学科研人员来说尤为重要。医学会议信息

已成为科研人员了解医学领域最新发展状况的重要情报源之一。会议信息主要分为会议预告信息和会议论文信息。

二、会议预告信息检索

参加学术会议有助于促进学术交流、共享科研成果、掌握专业发展动态,为进一步的专业研究和学术交流积累信息。及时获悉学术会议信息,即获取学术会议召开的时间、地点、主题和会议征文通知等内容,是撰写学术会议论文、参加学术会议的指南,通过网络形式发布会前信息已成为一种趋势。

(一) 国内医学会议预告信息

1. 中华医学会网站(<http://www.cma.org.cn>) 是中华医学会组织学术交流活动、开展继续医学教育的学术网站。通过浏览“会议通知”栏目,可获取由中华医学会各分会、中华医学系列期刊编辑部以及音像社主办的学术会议信息,包括会议介绍、日程、地点、征文以及会议网站链接等项内容。通过“学术月报”,按年度和会议类别浏览学术会议计划,包括主办方的分会名称、会议名称、重要内容和目的、会议时间和地点、参会人数、会期以及联系人和联系方式等项信息。注册会员还可以利用网站提供的“学术会议计划申报”“学术会议计划查询”栏目,发布和查询会议信息。

2. 医学会议在线(<http://www.medig.com.cn>) 汇集了大量的国内外医学会议信息。数据每日更新。设有最新医学会议、近期医学会议、医学会议报道、重要会议推荐、发布会议、会议课件、医学资讯等栏目。提供按科室归类会议信息的学科分类导航以及会议信息检索功能。用户可在首页左侧搜索框中键入检索词,选择会议类型(国际会议、国内会议、研讨会、讲习班)和会议地点,点击“搜索”,即可完成检索操作。网站还为注册会员提供发布会议信息、在线会议报名、在线提交会议论文等个性化服务,该网站同时还有会议报道、会议课件下载等内容。

3. 好医生会议(<http://conference.haoyisheng.com>) 好医生网站主要为医护人员提供业内信息及远程教育服务。其学术会议中心栏目提供近期会议预告、医学会议新闻、重点会议推荐展示、按学科浏览会议专题、会议现场视频直播、学术会议专家讲座视频、会议课件下载等服务。可按学科、按时间浏览或检索会议信息。会议信息主要包括会议名称、地点、日期、参会在线报名、征稿等项内容。注册用户可通过 E-mail 及时获得有关会议信息。

4. 中国学术会议在线(<http://www.meeting.edu.cn>) 是由教育部科技发展中心主办,面向广大科技人员的科学研究与学术交流信息服务平台。为用户提供境内会议预报、境外会议信息预报、会议在线报名、会议信息发布、名家讲堂、会议视频点播等服务。境内会议信息检索有会议分类检索和高级检索,同时还可在检索结果内检索。点击每条检索结果的会议信息都有会议基本信息、联系方式、会议背景介绍、会议征文范围及要求等。境外会议可以按分类浏览,每一类目录里又可按照国家浏览会议信息。

网站可以在线发布会议,点击主页“登录”框下方的“发布会议信息”,出现发布会议信息流程和会议在线报名流程,免费注册后按流程操作可发布会议信息或者在线报名参加学术会议。

(二) 国际医学会议预告信息

1. AllConferences(<http://www.allconferences.com>) 收集各种会议、展会的信息,并对收集的网页按主题分类,医学会议信息是该网站重点收录内容。用户可按照主题目录浏览,也可直接输入关键词快速检索或者选择高级检索,级检索可按名称(by Title)、城市(by City)、国家(by Country)、日期(by Date)、分类(by Category)、关键词(by Keyword)分别检索。检索结果有即将召开的会议信息,也有已经召开的会议信息。每条结果有会议名称、地点、时间、联系方式和会

议简介,有的还提供会议的 URL。通过主页上的 Submit a Conference,可以免费发布会议预告信息。

2. Medical Conferences (<http://www.medicalconferences.com>) 由英国医学会议公司 (Medical ConferencesComp) 创建,网站报道了近 100 个国家的约 7000 个即将举行的医学会议,每日更新。点击主页面上的“SEARCH”图标可进入会议检索页面,用户可使用关键词、会议名称、会议地点、及开会时间检索信息。如需会议详细信息,用户须免费注册,向会议组织者提交个人信息后,由会议的组织者将会议的详细信息传送到用户的电子邮箱。

3. Doctor's Guide to Internet (<http://www.docguide.com>) 简称 DGI, 是著名的美国医学信息检索站点。该网站的会议资源中心 (Congress Resource Centre, 简称 CRC), 专门收集正在召开的和即将召开的世界各国医学会议信息。可浏览、检索会议信息,还可通过“Submit your Congress/Conference”栏目,向 CRC 免费发布会议信息。会议资源中心提供三种检索方式:一是基本检索,在检索框中输入关键词或时间、地址名词,便可检索到相关的会议信息,如在检索框输入“China”,检索出即将在中国召开的会议信息。二是浏览检索,可按会议所属专业 (By Specialty)、会议日期 (By Date)、会议地点 (By Location) 逐层点击浏览会议信息。三是高级检索,可用逻辑运算符 AND、OR、NOT、短语检索、截词检索。在检索框内输入检索词或者检索表达式后进行检索,可获得所需要的会议名称、开会日期、会议地点、联系人、电话及 E-mail 地址等会议信息。此外,每一条的会议信息下都有会议召开地点的天气、航班、旅游、饮食、住宿、订票、货币汇率等信息。

4. 学术会议网医学会议预报 (<http://www.medical.theconferencewebsite.com>) 是专为用户提供免费查询国际上医学会议信息及继续医学教育课程的一站式网站,设有简单检索、高级检索两种途径。用户可在检索框中键入会议名称全称或缩写、会议地点、会议日期,查询包括会议的名称 (Conference Name)、地点 (Location)、时间 (Starts、Ends) 等信息。点击会议名称链接,可获得会议网站链接 (View conference website)、会议举办地地图链接 (View map of conference location) 等更为详细的会议信息。此外,用户可通过导航栏 (Navigation) 的“ADD A CONFERENCE”,发布会议信息。注册会员还可通过 E-mail,及时获得最新的会议信息。

5. HON 会议信息 (<http://debussy.hon.ch/cgi-bin/confevent>) 是健康在线基金会 (HON) 网站的栏目之一。HON 会议信息提供全球范围内医学学术会议和继续医学教育 (CME) 等信息 (<http://debussy.hon.ch/cgi-bin/confevent>)。可按会议时间 (Date)、关键词字顺 (Alphabet) 或学科分类 (Browse)、会议地点 (USA, Europe, Other) 浏览和检索会议。也可在检索框中键入检索词,在全文中检索会议信息。会议信息主要包括:会议名称、日期、地点、会议简介、关键词、会议类型、会议使用语言、会议网站链接等内容。用户可通过“Submit a conference”,发布会议信息;通过“Past conferences”,浏览已召开的会议信息;通过“Mailing service”,定制个性化服务。

三、国内会议论文检索

(一) 万方数据《中国学术会议论文库》

《中国学术会议论文库》收录了国家级学会、协会、研究会在国内组织召开的全国性学术会议论文,是国内收集学科最全、数量最多的会议论文数据库。它由中国科技信息研究所提供,由万方数据股份有限公司通过万方数据资源系统上网,向社会提供服务。《中国学术会议论文库》每年涉及 600 余个重要的学术会议,每年增补论文 15 000 余篇,专业范围包括自然科学、工程技术、农林、医学等领域。该数据库已作为万方知识服务平台的一部分,用户可在该平台跨库检索中选择“会议论文”检索该数据库全文资源,也可以利用学科分类浏览会议论文 (图 8-6)。



Figure 8-6 shows the 'Novelty Search' interface of Wanfang Data. It features a sidebar on the left for selecting document types, a central search area with various input fields and dropdown menus for search criteria, and a search button. The interface is designed for advanced searching across multiple databases.

图 8-6 中国学术会议论文库检索界面

(二) 中国知网《中国重要会议论文全文数据库》

中国知网数据库之一, 收录我国 1999 年以来中国科协及国家二级以上学会、协会、研究会、科研院所、政府举办的重要学术会议, 高校重要学术会议以及在国内召开的国际会议上发表的文献。用户可通过初级检索、高级检索、专业检索三种途径, 选择题名、主题、关键词、摘要、作者、会议名称、主办单位、会议地点等字段, 免费检索、浏览会议文献的题录和摘要等内容。此外, 系统还提供了会议主办单位导航系统, 用户可通过分类(学会、协会、研究会、联合会、研究中心)、地域两种方式直接进行浏览和检索。也可以直接在检索框中键入会议主办单位名称进行查找(图 8-7)。

Figure 8-7 displays the search interface of the CNKI 'China Important Conference Paper Full Text Database'. It includes a sidebar for document classification, a main search area with various input fields and dropdown menus for search criteria, and a search button. The interface is designed for advanced searching across multiple databases.

图 8-7 中国重要会议论文全文数据库

四、国外会议论文检索

(一) CPCI 国外会议文献检索系统

ISI Proceedings 原为美国科学情报研究所 (ISI) 创建的专门收录国际会议录的数据库, 包括 ISTP (Index to Scientific & Technical Proceedings, 科学技术会议录索引) 和 ISSHP (Index to Social Science & Humanities Proceedings, 社会科学及人文科学会议录索引) 两个版本。2008 年 10 月以后, ISI Proceedings 更名为 Conference Proceedings Citation Index (简称 CPCI, 会议论文

引文索引),并被整合到 Web of Science 检索平台。ISTP 更名为 CPCI-Science, ISSHP 更名为 CPCI-SSH。

CPCI 汇集了世界上最新出版的会议录资料,收录农业(agriculture)、环境科学(environmental sciences)、生物化学(biochemistry)、生物学(biology)、生物技术(biotechnology)、医学(medicine)、工程(engineering)、计算机科学(computer science)、化学(chemistry)、物理(physics)等自然科学与技术领域的著名国际会议文献。包括专著、丛书、预印本、期刊、报告等形式出版的国际会议论文文摘及参考文献索引信息,提供综合全面、多学科的会议论文资料,包括会议名称、主办机构、地点、论文篇名、论文摘要、参考文献等会议信息及会议文献信息,是科研人员了解和查找世界上权威会议文献最主要的检索工具。CPCI 为文摘索引型数据库,具有一般检索和高级检索两种途径以及全文链接功能,提供 1990 年以后出版的会议论文的文摘、出版信息、相关会议信息等内容,同时还收录 1999 年至今的文后参考文献,数据库每周更新,具体检索方法参见“Web of Knowledge”检索。

(二) OCLC ProceedingsFirst 国外会议论文检索系统

包括 PapersFirst(国际学术会议论文索引)和 ProceedingsFirst(国际学术会议录索引)两个会议论文数据库,收录了 1993 年以来世界范围的研讨会、专题会、学术报告会、座谈会、博览会等各种会议的论文题录信息,包括英国不列颠图书馆文献提供中心(British Library Document Supply Center,简称 BLDSC)出版的会议论文及资料。PapersFirst 收录会议论文 650 万余条记录。ProceedingsFirst 收录会议录 19 万 2 千余条记录。两个数据库的检索途径及字段基本相同,均提供基本检索、高级检索和专家检索等途径,关键词、作者、会议名称、会议地址、会议日期等 27 个检索字段。会议论文题录主要包括作者、论文题名、资料来源、语种、会议名称等项内容。数据每月更新两次。

(仇晓春)

第三节 专利文献检索

一、概述

(一) 专利

1. 专利(patent) 是专利权的简称,指一项发明创造向国家专利局提出专利申请,依法审查合格后,向专利申请人授予的在规定时间内对这项发明创造享有的专有权。

各个国家都采用建立专利制度的形式来保护专利权。专利制度的核心是专利法,它主要从专利权的授予和保护两个方面对专利事务进行规范。我国于 1984 年颁布并于同年 4 月 1 日正式实施了《中华人民共和国专利法》。

2. 专利的属性 专利权属于知识产权,具有独占性、地域性和时间性。

独占性也称专有性或排他性。专利权人对他的发明创造享有的独占性的制造、使用、销售和进口等实施权。

地域性指一个国家授予的专利权,只在本国有效,对其他国家没有约束力,任何国家都没有保护别国专利的义务。

时间性指任何专利的保护有一定的法律期限。专利权人对其发明创造所拥有法律赋予的专有权只在法律规定的期限内有效。多数国家自专利申请日之日起,发明专利的保护期限是 20 年,实用新型和外观设计专利是 10 年,药品专利可适当延长。美国规定 FDA 批准的药品专利保护可延长 5 年,但不超过产品上市之日后的 14 年。欧共体规定自 1993 年 1 月 2 日以来,药品专利在获得有关卫生部门的生产许可后,如果专利保护期不足 15 年,可以延长 5 年。日本与



美国类似,最多可延长5年。

3. 专利的类型 具体划分各国不尽相同,我国分为发明专利、实用新型专利和外观设计专利三种类型。

(1) 发明专利:是对产品、方法或者其改进所提出的新的技术方案。

(2) 实用新型专利:是对产品的形状、构造或者其结合所提出的适于实用的新的技术方案。

(3) 外观设计专利:是对产品的形状、图案、色彩或者其结合所做出的富有美感并适于工业上应用的新设计。

4. 专利优先权(patent priority) 是指专利申请人就其发明创造第一次在某国提出专利申请后,在法定期限内,又就相同主题的发明创造提出专利申请的,根据有关法律规定,其后在他国申请仍以第一次专利申请的日期作为其申请日,专利申请人依法享有的这种权利,就是优先权。专利优先权的目的在于排除在其他国家抄袭此专利者,有抢先提出申请,取得授权的可能。

5. 同族专利(patent family) 具有共同优先权的由不同国家或国际专利组织多次申请、多次公布或批准的内容相同或基本相同的一组专利。由至少一个共同优先权联系的一组专利,称一个专利族(patent family)。在同一专利族中每件专利文献被称作专利族成员,同一专利族中每件专利互为同族专利。在同一专利族中最早优先权的专利称基本专利。

6. 失效专利(invalid patent) 下列三种情况之一的为失效专利:①超过专利法定保护期限;②专利权人以书面声明放弃其专利权;③未按规定缴纳年费。失效专利的技术可以无偿使用,是促进技术创新、开发新产品的宝贵资源。

(二) 专利文献

专利文献是包含已经申请或被确认为发现、发明、实用新型和工业品外观设计的研究、设计、开发和试验成果的有关资料,以及保护发明人、专利所有人及工业品外观设计和实用新型注册证书持有人权利的有关资料的已出版或未出版的文件(或其摘要)的总称。

专利文献主要指记载发明创造内容的技术文献,是技术创新信息的载体。主要有专利公报和专利说明书,其次是题录、文摘、索引、分类表等。据世界知识产权组织统计,世界上90%~95%以上的发明创造成果都出现在专利文献中,其中70%不出现在非专利文献中。充分利用专利信息可以节约60%的研究时间和40%的科研经费。专利信息检索和分析不仅可以避免重复研究和专利侵权,同时可以有效地保护自己的科研成果。因此,对于专利文献的检索、分析和利用,已成为科学研究与技术创新过程中不可缺少的重要组成部分。

(三) 专利分类法

目前全球主要的专利分类体系包括国际专利分类(IPC),欧洲专利分类(ECLA、ICO),日本专利分类(FI、F-Term),美国专利分类(USPC),联合专利分类(CPC)。

1. 国际专利分类(International Patent Classification, IPC) 是国际专利组织(WIPO)管理的分类体系,是目前许多国家普遍采用的专利文献分类工具。IPC是其他分类体系细分的基础,但是IPC分类有的分类不够细,分类条目较宽等不足。

IPC分类主要是对发明专利,包括出版的发明专利申请书、发明证书说明书、实用新型说明书和实用证书说明书等(统称为专利文献)进行分类。IPC按五级分类:部、大类、小类、主组、分组。部以下的分类会阶段性调整、增加。

一个完整的IPC分类号由代表部(1个字母)、大类(2个数字)、小类(1个字母)、大组(1~3个数字)或小组(2~4个数字)的符号构成。例如:A01B 1/02。

2. 欧洲专利分类(ECLA、ICO) 其中ECLA是欧洲专利局在IPC基础上的进一步细分得到的分类体系,ICO是针对ECLA的标引系统,是用以标引附加信息的分类体系。

3. 日本专利分类(FI、F-Term) FI分类是基于IPC分类的细分类,F-Term是专门用于计算机检索的分类体系。

4. 美国专利分类(USPC) 该分类体系不以 IPC 为基础,其分类体系比较独特,检索的文献也限于美国专利文献。

5. 联合专利分类(Cooperative Patent Classification, CPC) 由于上述多个分类体系的存在引起了对专利文献的分类标准的不统一,只有熟悉各个分类体系,才能对专利文献进行有效的检索,这增加了检索的成本,为了在世界范围内建立一个统一的分类体系,并继承已有的分类体系的优点,欧洲专利局和美国专利商标局在 2010 年 10 月 25 日,签署了合作文件,同意在 IPC 分类基础上共同合作,创建并实施 CPC,并于 2012 年 10 月 1 日发布了 CPC 架构。

CPC 的架构源于 ECLA 分类系统,编码规则与 IPC 类似,其保留原 IPC 分类号,并将原 ECLA、ICO 中“/”之后的字母数字部分全部改为数字形式。

(四) 专利书目数据特征

专利文献著录项目所代表的信息包括:技术信息、法律信息和文献外在形式信息。

专利的技术信息是通过专利文件中的说明书、附图等文件部分详细展示出来的。便于人们从各种角度便捷地了解该发明创造信息,如发明创造名称、专利分类号、摘要等技术信息。

专利法律信息包括专利保护的范围,专利的权利人、发明人,专利的生效时间,专利申请的标志等。通过专利分类号(专利保护的范围)、申请人、发明人、专利权人、专利申请号、申请日期、优先申请号、优先申请日期、优先申请国家、文献号、专利或专利申请的公布日期、国内相关申请数据等专利文献著录项目来揭示不同法律信息特征。

专利文献外在形式的信息著录项目主要是文献种类的名称、公布专利文献的国家机构、文献号、专利或专利申请的公布日期。

专利文献作为一种专类文献,除了标题(专利名称)、作者(专利权人、申请人等)、摘要等基本要素外,尚有一些特殊的外部特征和内容特征。下面以我国的专利文献为例予以说明。

1. 专利申请号 指专利局受理发明创造专利申请时所给予的顺序号。

2. 专利号 指专利申请授权后该专利的号码。

3. 专利申请人(applicants) 可以是该发明创造的发明人(inventors)、设计人(非职务发明创造)或其所属单位(职务发明创造),也可以是该发明创造的合法受让人或继承人,或者与中国签订协议或与中国共同参加国际条约或按对等原则办理的国家的外国人、外国企业或外国其他组织。

4. 专利权人(patentee) 是专利权的所有人及持有人的统称。即专利申请被批准时,被授予专利权的专利申请人。专利权人既可以是单位也可以是个人。

独占实施权(独占制造权、独占使用权、独占销售权)是专利权人最基本的权利。专利权人可以通过订立专利实施许可合同许可其他人使用其专利技术;专利权人还可以向其他人转让其专利技术;专利权人应当按时向专利局交纳年费,以维持其专利权的有效性;当有人提出宣告专利权无效的请求时,专利权人将作为被请求人出庭应诉;当专利局颁发强制许可或政府有关部门发放“计划许可”时,专利权人也将作为当事人参与该法律关系等。

5. 专利受让人(assignee) 是指被专利权人让与专利权的所有人的统称,专利受让人既可以是单位也可以是个人。

二、国内专利检索

(一) 国家知识产权局专利检索系统

由中华人民共和国国家知识产权局(SIPO)建立的“专利检索与服务系统”(http://www.pss-system.gov.cn/)向公众提供专利检索和专利分析服务。

可检索 103 个国家、地区和组织的专利数据,其中涵盖了美国、日本、韩国、英国、法国、德国、瑞士、俄罗斯、欧洲专利局和世界知识产权组织。公众可免费获取专利全文图像文



件,注册用户还可使用专利分析服务。

该检索系统支持以下检索规则:①逻辑运算符: and、or、not;②支持截词符(半角字符): #、+、?、# 表示一个强制存在的字符,+ 表示任何长度的字符串,? 表示 1 个或没有字符;③支持词组检索:用英文双引号“ ”;④可输入多个关键词,中间用空格分隔,空格表示逻辑“或”、“or”的关系;⑤申请号格式:文献的申请国+申请流水号;⑥公开(公告)号格式:文献的公开国+公开流水号+公布级别。

1. 专利检索 主要有常规检索、表格检索、多功能查询。

(1) 常规检索:提供申请要素(包括摘要、关键词、权利要求和分类号)、申请号、公开(公告)号、申请(专利权)人、发明人和发明名称 6 个检索入口(图 8-8)。

图 8-8 SIPO 常规检索界面

(2) 表格检索:表格检索是根据数据库可检索字段按照表格方式设置的检索功能(图 8-9),在各字段检索框中可输入单个检索词或多个检索词组成的检索式,系统默认各检索字段为逻辑“与”的关系。默认为中外专利联合检索。中文专利检索可限定专利文献类型(公开文献、授权公开文献)和发明类型(发明、实用新型、外观设计)。

图 8-9 SIPO 表格检索

如:检索 2005 年至 2014 年公开的关于医用生物传感器方面的专利,操作步骤为(图 8-10):①选择“中外专利联合检索”;②在“公开(告)日”字段检索框中输入:2005 2014。时间段语法选择“:”;③在“发明名称”字段输入:生物传感器;④在“说明书”字段输入:医学 医疗 医用;⑤点击【检索】按钮,完成检索,结果如图 8-10 所示。

检索完成后,系统以摘要的形式显示检出专利的基本信息,通过页面提供的功能按钮可进一步对检出结果进行处理,包括查看文献详细信息(可查看专利说明书全文)、同族信息、引证/被引证信息、法律状态、申请人(专利权)基本信息等。

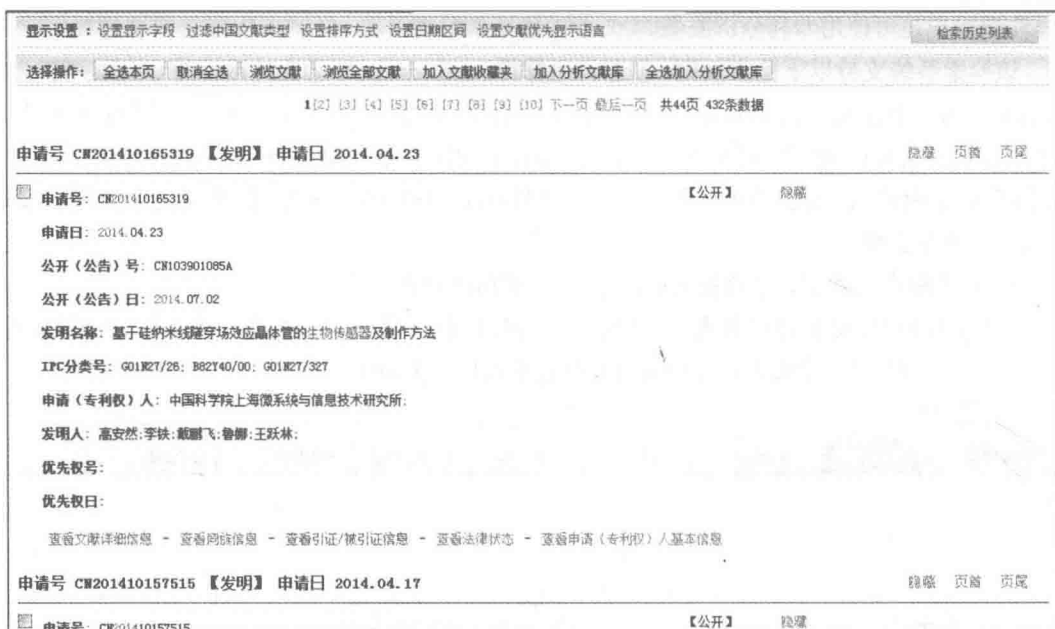


图 8-10 SIPO 专利检索结果

(3) 其他查询: 包括 IPC 分类、同族专利、法律状态、引证 / 被引证等查询等, 通过多功能查询器实现。

IPC 分类表查询: 提供分类浏览、中文含义和英文含义查询, 如图 8-11。

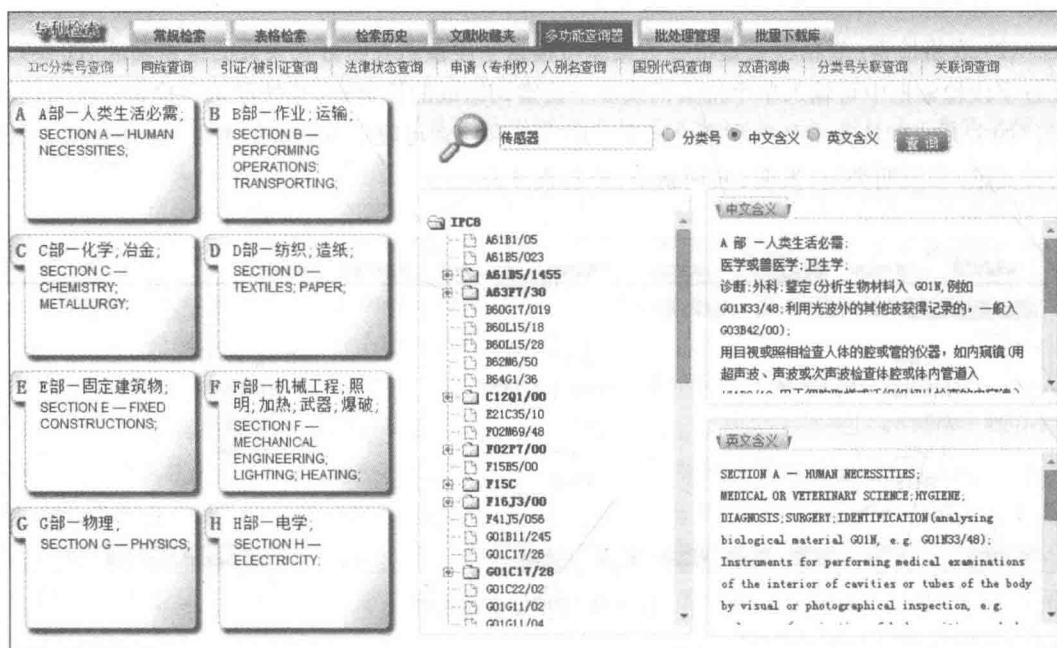


图 8-11 SIPO-IPC 分类号查询

分类号关联查询: 世界各国使用的专利分类体系不尽相同, 世界知识产权组织 (WIPO) 使用的是国际专利分类体系 (IPC), 美国使用的主要是美国专利分类体系 (USPC), 欧洲专利局 (EPO) 使用的主要是基于 IPC 的欧洲专利分类体系 (ECLA/ICO), 日本使用的主要是基于 IPC 的日本专利分类体系 (FI/FT), 韩国和中国使用的均为国际专利分类体系 (IPC)。为便于对照查询不同的专利分类号, 系统提供了 IPC 分类号对应法 ECLA、UC、FI、FT 的关联查询。

关键词查询：系统提供基本词所属 IPC 分类号及英文解释，以及与基本词相关联的上位词、下位词、同义词、近义词、错位词、反义词以及族首词等关联词信息，图 8-12。

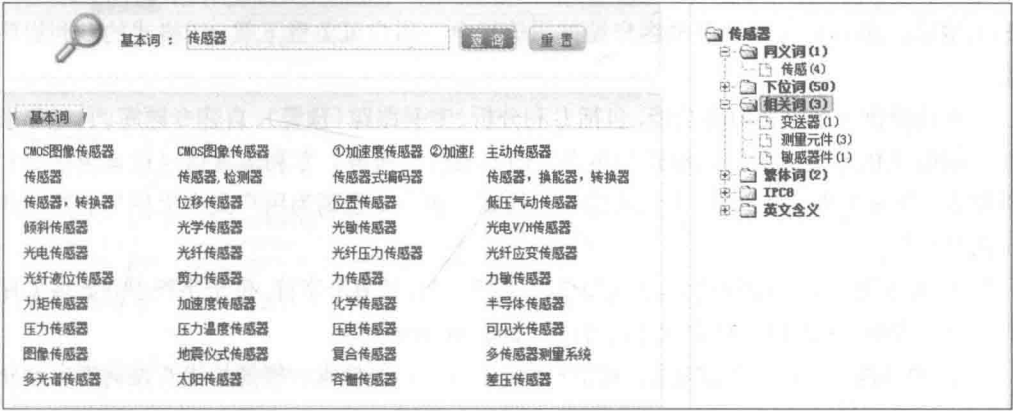


图 8-12 SIPO- 关联词查询结果

2. 专利分析 SIPO 提供的专利分析包括快速分析、定制分析和高级分析，每个分析模块下面提供多种分析功能(表 8-1)。主要从区域、技术领域、申请人、发明人角度分析其趋势、分布、构成等，并以图表的形式直观显示。每一个功能模块下还可进一步分析，如“快速分析”的“区域分析”进一步还可进行区域构成分析、区域趋势分析、区域技术领域分析、区域申请人分析、区域发明人分析等(图 8-13)。

表 8-1 SIPO 专利分析模块及功能

分析模块	分析项目
快速分析	区域分析 技术领域分析 申请人分析 发明人分析 中国专项分析
定制分析	技术衍变趋势 地域性信息分布 核心申请人统计 核心发明人统计 核心技术统计 核心引用文献分析 企业发展定位分析
高级分析	列表分析 矩阵分析

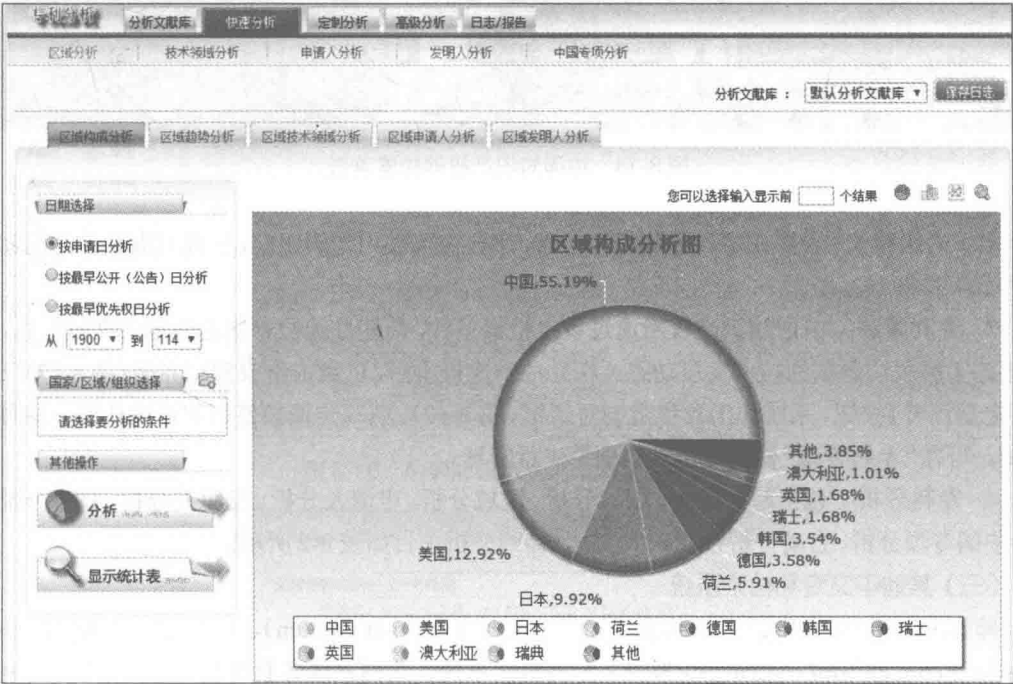


图 8-13 SIPO 快速分析

(二) 中国知识产权网

由知识产权出版社有限责任公司建立的专利检索平台 (<http://search.cnipr.com/>), 提供对中国专利和国外(美国、日本、英国、德国、法国、加拿大、EPO、WIPO、瑞士等 98 个国家和组织) 专利的检索。用户还可通过手机客户端使用该平台。用户可免费下载 .tif 格式的专利说明书全文。

该系统提供了个性化服务功能, 包括专利分析、专利跟踪(预警)、自建专题库、自建导航检索等。同时提供专利运营信息检索和热点专题专利信息服务。专利运营信息检索包括专利权转移检索、专利质押保全检索、专利实施许可检索。热点专题则为用户集中提供当前热点研究的国内外专利。

1. 检索规则 支持模糊检索, ? 代替单个字符, % 代替多个字符, 位于字符串开始或末尾时模糊字符可省略; 字段内各检索词之间可进行 and、or、not 运算。

2. 检索方式 提供一框式检索、智能检索、表格式高级检索。智能检索实现词汇自动转换(同义词、近义词等)。

(1) 一框式检索: 系统默认检索界面, 类似于 google 样检索(图 8-14), 键入任何一个检索词后, 点击【检索】或【智能检索】按钮即可获得检索结果(图 8-15)。



图 8-14 中国知识产权网检索首页

(2) 高级检索: 表格式字段限制组合检索, 字段内可使用逻辑组配, 字段间默认的逻辑组配为“and”(图 8-16)。

3. 专利跟踪 中国知识产权网通过设定检索条件, 周期性地检索指定领域的文献信息, 从而跟踪了解关注领域的技术发展动态。在完成一次检索后, 可点击结果预览界面(图 8-17)下方的【定期预警】按钮, 系统弹出预警定制对话框(图 8-17), 填写定期预警名称后确认即完成预警定制。可在“专利管理”模块对定制的预警进行管理。

4. 专利分析 包括趋势分析、国省分析、区域分析、申请人分析、发明人分析、技术分类分析、中国专项分析、自动分析报告八大类, 具体的分析项目如表 8-2 所示。

(三) 其他中文专利检索系统

除 SIPO 外, 还有中国专利信息中心 (<http://www.cnpat.com.cn/>)、专利之星 (<http://search.patentstar.com.cn/frmLogin.aspx>) 等中文专利检索系统。这些系统各具特色, 收录的中国专利均始于 1985 年, 用户需付费才可获取专利全文。



关键词 二次检索 过滤检索 重新检索

检索分类 全部 (2767) 中国发明专利 (2212) 中国实用新型 (555) 更多 >>

☐ 全选 ☒ 默认排序 ☐ 公开日 ☐ 申请日 ☐ 显示10条

☐ 一种集中式Mesh无线传感器网络冗余路径调度及管理方法 发明专利 有效

申请号: CN200910104021.7 申请日: 2009.06.05
 公开(公告)号: CN101605372A 公开(公告)日: 2009.12.16
 申请(专利权)人: 重庆邮电大学
 分类号: H04W40/02 (2009.01)I; H04W64/18 (2009.01)I
 优先权:

摘要: 本发明请求保护一种集中式Mesh无线传感器网络冗余路径调度及管理方法, 涉及无线传感器网络。无线传感器网络中为避免因一个中继节点的故障影响其他设备的通信, 本发明通过建立网络设备关联矩阵来反映设备节点间的连接状态, 并以路由路径节点序列中相邻节点为三角形两顶点则寻找该相邻节点的替代节点序列, 选择出合适的替代路由节点组成一条冗余路径。系统管理器针对路由路径自动地建立并维护相应的冗余路径, 保证节点在源路由路径不能正常通信的情况下可以及时切换到冗余路径保持通信, 从而提高数据通信的可靠性。冗余路径与相应的路由路径不会有链路上的重合。能广泛应用于无线传感器网络。

☐ 用于使用磁感应无线传感器测量物体的形状的系统和方法 发明专利 无效

申请号: CN201010600615.5 申请日: 2010.12.20
 公开(公告)号: CN102166120A 公开(公告)日: 2011.08.31
 申请(专利权)人: HXP股份有限公司
 分类号: A61B5/103 (2006.01)I
 优先权: 2009.12.21 US 12/643,995

摘要: 一种用于使用磁感应无线传感器测量物体的形状的系统和方法, 包括: 利用磁感应无线传感器的磁环天线至少部分地包围物体, 其中磁环天线感应取决于物体的形状; 以及在与磁环天线耦合的天线匹配电路处响应于磁环天线的电感提供特定电容, 以使得磁环天线和天线匹配电路构成谐振电路, 并且该谐振电路具有固定的谐振频率, 其中, 所述特定电容用于测量物体的形状。

★收藏 ...

图 8-15 中国知识产权网检索结果预览页面

☐ 全选 ☐ 数据范围 ☐ 同义词 ☐ 保存表达式

☐ 中国 ☐ 中国发明申请 ☐ 中国实用新型 ☐ 中国外观设计 ☐ 中国发明专利 ☐ 台湾 ☐ 香港

☐ 主要国家和地区 ☐ 美国 ☐ EPO ☐ 日本 ☐ WIPO ☐ 英国 ☐ 瑞士 ☐ 德国 ☐ 韩国 ☐ 法国 ☐ 俄罗斯

☐ 其他国家和地区 ☐ 东南亚 ☐ 奥地利 ☐ 阿拉伯 ☐ 意大利 ☐ 意大利 ☐ 非洲地区 ☐ 加拿大 ☐ 瑞典 ☐ 西班牙 ☐ 更多

申请(专利)号: 例如: CN20144866.5 申请日: 例如: 20101010
 公开(公告)号: 例如: CN1397751 公开日: 例如: 20110105
 名称: 例如: 计算机 摘要: 例如: 计算机
 权利要求书: 例如: 计算机 说明书: 例如: 计算机
 申请(专利权)人: 例如: 华为 发明(设计)人: 例如: 滕学平
 国际专利主分类号: 例如: G06F15/16 国际专利分类号: 例如: G06F15/16
 地址: 例如: 北京市海淀区 国家代码: 例如: 北京
 同族专利: 例如: US24512305F 优先权: 例如: CN20112980.2
 代理机构: 例如: 舜光 代理人: 例如: 里内南

and or not () >>
 and
 生成表达式

编号	表达式名称	检索数据库	命中数量	检索时间	操作
20	第 1 页 共 1 页				显示0到0, 共0记录

图 8-16 中国知识产权网检索高级检索

新建定期预警

定期预警名称: 例如: 医用传感器

表达式: 名称=(传感器) and 摘要=(医用 or 医学 or 医疗)

图 8-17 中国知识产权网检索预警定制对话框

表 8-2 中国知识产权网专利检索系统分析功能

分析模块	分析项目
趋势分析	总体趋势分析
区域分析	区域趋势分析、区域构成分析、区域技术分类构成、区域申请人构成
申请人分析	申请人趋势分析、申请人构成分析、申请人区域构成、申请人技术分类构成、申请人综合比较、合作申请人分析
发明人分析	发明人趋势分析、发明人构成分析、发明人技术分析、合作发明人分析
技术分类分析	技术分类趋势分析、技术分类构成分析、技术分类区域构成、技术分类申请人构成、技术关联度分析
中国专项分析	专利类型分析、国省分布状况
高级分析	全文聚类分析、引证分析
自动分析报告	总体发展趋势、其他专项分析、专利申请区域分析、主要技术领域分析、主要竞争者分析、主要发明人分析

三、国外专利检索

(一) 国外专利资源分布概述

因特网上的国外专利数据库很多,由各国专利局官方网站(如 SIPO、USPTO)、公益性免费专利数据库网站(如 free patents online)、商业性专利数据库网站(如 Delphion)、大型国际商业性联机检索系统(如 STN、DIALOG、ISI)等多种组织或机构提供。

由各国专利局官方或国际组织提供的大多可免费获得相关资料。其中专利容量大、检索功能强的有美国专利数据库、欧洲专利局专利数据库、世界知识产权局 PCT 电子公报、加拿大知识产权局专利数据库和日本工业产权数字图书馆专利数据库,都可以免费下载专利说明书。德温特创新索引数据库(Derwent Innovations Index, 简称 DII)收录的专利数量最大,涉及的国家 and 地区最多,但不能免费检索。大型国际商业联机检索机构所提供的专利检索多以 DII 为基础。

(二) 美国专利商标局专利检索

该检索系统提供申请专利(AppFT: Applications)和授权专利(PatFT: Patent)两种检索入口(<http://patft.uspto.gov/>, 图 8-18),并提供专利分类表、专利法律状态、专利转让数据库等相关数据库的检索。

提供快速检索、高级检索和号码检索。1790—1975 年之间的专利只能通过专利号和美国专利分类查找。

该系统免费提供 1976 年以来授权的专利说明书和 2001 年 3 月以后的专利申请文件,其专利文献全文有文本和图像两种形式。文本形式无图片,图像为原件扫描而成的 TIFF 文件,需要下载安装 TIFF 图片插件方可阅读。

(三) 欧洲专利局

欧洲专利组织通过 Espacenet 免费提供专利检索服务。该系统有多个登录入口,登录其引导页面(<http://www.epo.org/searching/free/espacenet.html>),包括欧洲专利组织(EPO)、欧洲委员会(EC)以及欧盟成员国的 Espacenet 站点。选择任何一个入口都可利用该检索系统世界专利数据库(worldwide)、欧洲专利局数据库(EP database)和世界知识产权组织数据库(WIPO database)。

1. 检索规则 支持布尔逻辑 and、or、not; 词组需要加双引号“, 否则系统按逻辑“与”处理,如 lung cancer 相当于 lung and cancer。支持*、?、#截词检索。

2. 检索模式 Espacenet 提供智能检索、高级检索和分类检索(Classification search)3 种检索模式。专利说明书为 PDF 格式,但不允许全部下载,只能单页下载和打印。

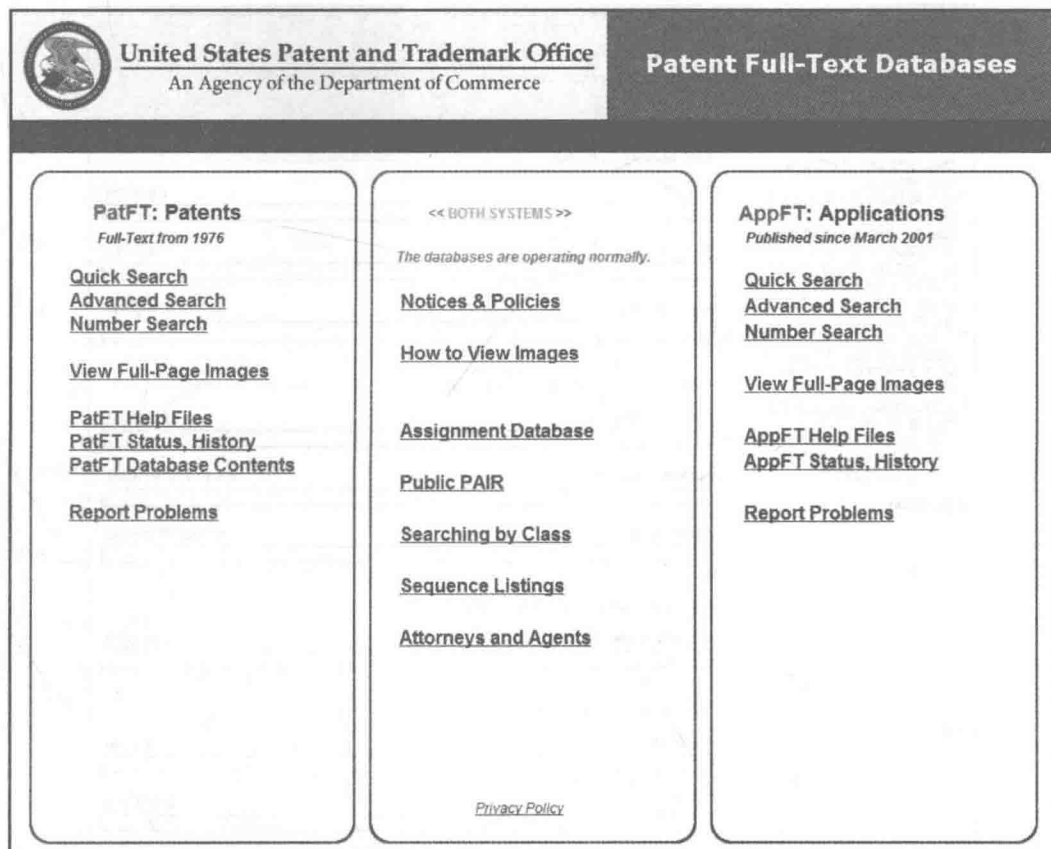


图 8-18 USPTO 专利检索主页

(1) 智能检索: 在检索框中输入任意词、词组、数字, 系统会自动转换或在相应的字段中去检索。如输入数字, 系统会自动将其当着专利号、申请号、优先权号、公开号、批准号的一部分去检索。

(2) 高级检索: 为填表式检索, 首先需要选择数据库(EP、Worldwide、WIPO 任选一个), 然后在相应的字段框中输入符合格式的检索提问, 点击“Search”按钮, 完成检索(图 8-19)。

(3) 分类检索: 提供基于 IPC 的联合专利分类(CPC)检索, 方便专利审查员和用户在进行专利检索时访问同一个类别的专利文档集合, 以实现分类资源共享。

3. 检索结果 包括专利基本信息(Bibliographic data)、专利描述(Description)、专利权(Claims)、专利缩图(Mosaics)、专利原文(Original document)、被引文档(Cited documents)、引证文档(Citing documents)、法律状态(INPADOC legal status)、同族专利(INPADOC patent family)等 9 个方面的信息(图 8-20)。

(四) 德温特创新索引数据库

德温特创新索引数据库整合了 Derwent 最著名的 World Patent Index(世界专利索引)和 Patent Citation Index(专利引文索引), 收录来自世界 40 多个专利机构的 3 千多万项专利, 数据可回溯至 1963 年。分为化学(Chemical Section)、电子电气(Electrical and Electronic Section)和工程技术(Engineering Section)三部分, 通过 web of science 平台综合检索世界各国的专利文献(图 8-21)。

较之其他专利检索系统, 其收录范围更广、数据量大, 但并不能提供全部专利的全文。此外, 还加入了 WoS 平台其他数据库的一些元素, 如“被引专利检索”(Cited patents search)、“化合物检索”(Compound search), 检索结果可对学科分类、专利权人、发明人、IPC 代码、德温特分类代码、德温特手工代码等进一步提炼, 并对检索结果进行计量分析。

Smart search Advanced search Classification search	Advanced search Select the collection you want to search in EP - complete collection including full text of European published applications Enter your search terms - CTRL-ENTER expands the field you are in Enter keywords in English Title: e.g. motor Title or abstract: e.g. hair Keyword(s) in full text: e.g. hair Enter numbers with or without country code Publication number: e.g. EP1883031 Application number: e.g. EP20070010825 Priority number: e.g. DE20021036409 Enter one or more dates or date ranges Publication date: e.g. 20070919 Enter name of one or more persons/organisations Applicant(s): e.g. IBM Inventor(s): e.g. Siemens Enter one or more classification symbols IPC: e.g. H02M7/537 H03K17/687
Quick help + How many search terms can I enter per field? + How do I enter words from the title or abstract? + How do I enter words from the description or claims? + Can I use truncation/wildcards? + How do I enter publication, application, priority and NPL reference numbers? + How do I enter the names of persons and organisations? + What is the difference between the IPC and the CPC? + What formats can I use for the publication date? + How do I enter a date range for a publication date search? + Can I save my query?	
Related links +	

图 8-19 Espacenet 高级检索界面

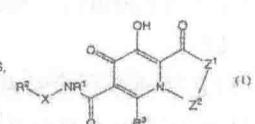
Search Results list My patents list (0) Query history Settings Help Refine search → Results page 1 → US2014200209 (A1)	Bibliographic data: US2014200209 (A1) — 2014-07-17 ★ In my patents list Previous 1/500 Next EP Registrar Report data error Print
US2014200209 (A1) Bibliographic data Description Claims Mosaics Original document Cited documents Citing documents INPADOC legal status INPADOC patent family	POLYCYCLIC CARBAMOYL PYRIDONE DERIVATIVE HAVING HIV INTEGRASE INHIBITORY ACTIVITY Page bookmark US2014200209 (A1) - POLYCYCLIC CARBAMOYL PYRIDONE DERIVATIVE HAVING HIV INTEGRASE INHIBITORY ACTIVITY Inventor(s): JOHNS BRIAN A [US]; KAWASUJI TAKASHI [JP]; TAISHI TERUHIKO [JP]; TAODA YOSHIYUKI [JP] ± Applicant(s): VIV HEALTHCARE COMPANY [US]; SHIONOGI & CO [JP] ± Classification: - international: C07D471/04; C07D471/14; C07D471/22; C07D498/14; C07D498/20; C07D498/22 - cooperative: C07D471/04; C07D471/14; C07D471/22; C07D498/04; C07D498/14; C07D498/20; C07D498/22 Application number: US201414211364 20140314 Priority number(s): US201414211364 20140314; JP20050131161 20050428; JP20050312076 20051027; US201313763174 20130208; US201213352686 20120118; US20060919386 20060428; WO2006US16604 20060428 Also published as: WO2006116764 (A1); US2013172559 (A1); US8778943 (B2); US2012115875 (A1); US8410103 (B2); more
Quick help + What does A1, A2, A3 and B stand for after a European publication number? + What happens if I click on "In my patents list"? + What happens if I click on the "Register" button? + Why are some sidebar options deactivated for certain documents? + How can I bookmark this page? + Why does a list of documents with the heading "Also published as" sometimes appear, and what are these documents? + Why do I sometimes find the abstract of a corresponding document? + What happens if I click on the red "patent translate" button?	Abstract of US2014200209 (A1) Translate this text into Albanian Google Translate powered by Google The present invention is to provide a novel compound (I) shown below, having the anti-virus activity, particularly the HIV integrase inhibitory activity, and a drug containing the same, particularly an anti-HIV drug, as well as a process and an intermediate thereof. (wherein Z1 is NR4; R1 is hydrogen or lower alkyl; X is a single bond, a hetero atom group selected from O, S, SO, SO2 and NH, or lower alkylene or lower alkenylene in which the hetero atom group may intervene; R2 is optionally substituted aryl; R3 is hydrogen, a halogen, hydroxy, optionally substituted lower alkyl etc.; and R4 and Z2 part taken together forms a ring, to form a polycyclic compound, including e.g., a tricyclic or tetracyclic compound.) 
Sitemap Accessibility Legal notice Terms of use Last updated: 11.06.2014 Worldwide Database 5.8.17, 92p	

图 8-20 Espacenet 专利详细信息



图 8-21 DII 专利检索主页面

(五) 免费专利在线

免费专利在线 (free patents online, <http://www.freepatentsonline.com/>) 目前提供美国专利、美国专利申请、欧洲专利、日本专利文摘、WIPO (世界知识产权组织) 专利和德国专利的查询与下载 (图 8-22)。

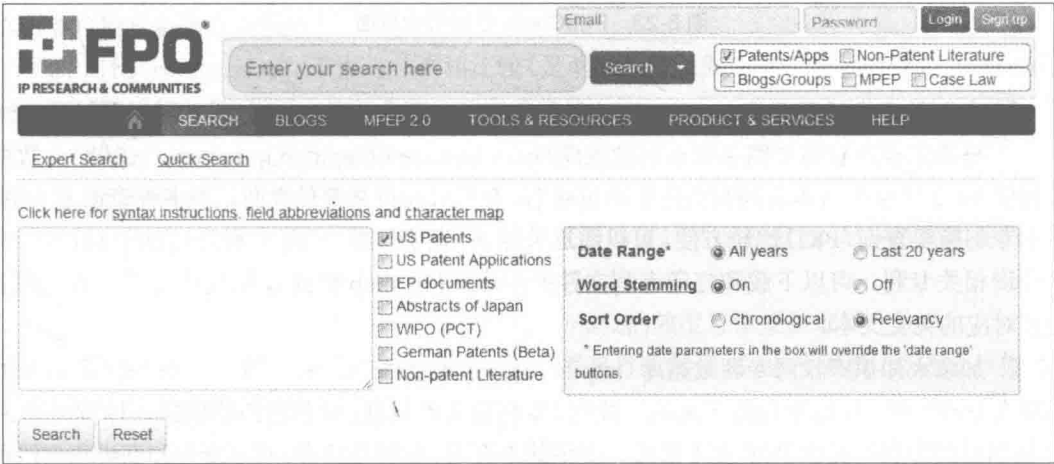


图 8-22 免费专利在线检索主页面

系统提供快速检索、专家检索 (Expert Search)。

快速检索支持表单式构建逻辑组配检索, 可从专利号、创作人、登记日、创作人所在城市、题名等 20 多个方面进行组合检索。

注册用户 (免费注册) 可在线保存检索历史、创建文件夹、获取相关新专利出现时的提醒等其他服务; 支持 RSS 订阅, 所有专利按学科分为近千个类提供 RSS, 每周更新, 用户可以订阅感兴趣内容的 RSS, 以便能及时跟踪新的专利。

(六) 世界知识产权局专利检索系统

世界知识产权局(WIPO)的专利检索(<http://www.wipo.int/pctdb/en/>)提供专利合作条约(Patent Cooperation Treaty, PCT)成员的国际专利申请查询。检索方式包括: 字段组合检索(Field Combination)、简单检索、高级检索、跨语言扩展检索(Cross Lingual Expansion)和浏览。

(七) PubChem 化合物专利检索

由美国国立医学图书馆提供的化合物专利检索, 包括化学药品的专利, 其检索除了常规的主题检索外, 尚可通过化学结构(structure)和分子式(molecular formula)检索(图 8-23)。

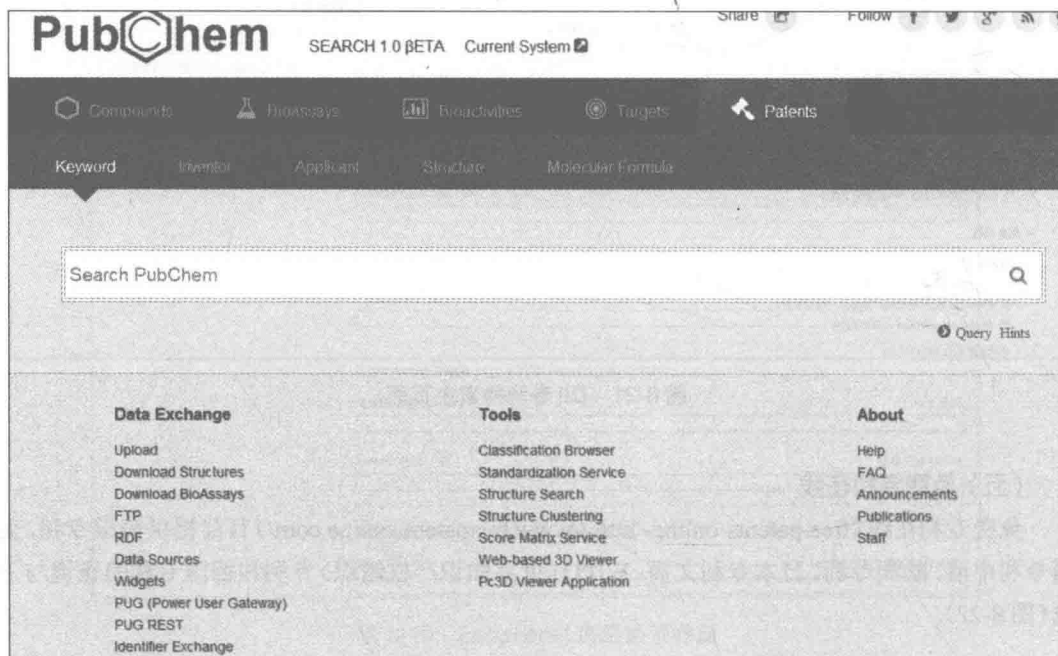


图 8-23 PubChem 专利检索页面

(八) 其他专利检索系统

1. 日本工业产权数字图书馆专利数据库(http://www.ipdl.inpit.go.jp/homepg_e.ipdl) 数据库提供 1922 年至今日本专利和实用新型说明书, 有日文和英文两种界面。在多种查询方式中, 日本专利摘要查询(PKI)比较方便, 可以通过关键词、发明日期、国际专利分类(IPC)以及专利号查询相关专利。可以下载和打印专利文件的图像文件(.bmp 格式), 年代较近的专利说明书还有对应的英文文本。

2. 加拿大知识产权局专利数据库(<http://brevets-patents.ic.gc.ca>) 该库包括 1920 年以来的加拿大专利文献, 包括专利的著录项目数据、专利的文本信息、专利的扫描图像。1978 年 8 月 15 日以后的专利可提供 PDF 文本形式。提供基本查询、专利号检索、布尔查询和高级查询四种形式。该数据库输出结果采用网页的形式, 主页面内容为专利名称、摘要、发明者、申请和授权时间、优先权等基本信息, 可以通过链接选择显示或下载专利申请书的各项内容, 使用非常简便, 另外该数据库还可以查询到该专利授权的当前状态。由于欧洲专利组织 Spacenet 的世界专利库中, 加拿大专利数据始于 1970 年, 因此, 要系统地查询加拿大专利, 应首选加拿大知识产权局专利数据库。

(赵文龙)



第四节 循证医学资源检索

一、概 述

(一) 循证医学

循证医学(evidence-based medicine, EBM)意为“遵循证据的医学”,是一种积极寻求和应用最佳证据,用以指导临床实践的医学。是建立在现代临床流行病学和临床研究基础上的临床医学,与传统经验医学最重要的区别是结果的可重复性强。循证医学的主要内容包括“四大原则”、“三大要素”和“两大核心”。

“四大原则”:①基于问题(临床关注的问题或重大的科学问题)的研究;②参考当前最好的证据决策;③关注实践的效果及后效评价;④止于至善。

“三大要素”:①参考当前所能得到的最好的临床研究证据:医疗决策(即患者的处理、治疗指南和医疗政策的制定等)应在现有的最好的临床研究依据基础上作出,要求自觉、准确、公正地根据现有最好的证据来决定对每一位患者的治疗选择;②参照医师自己多年临床实践经验和在检查患者所得到的第一手临床资料;③尊重患者的选择和意愿:将患者的意愿提到很高的程度上,如癌症患者要考虑到其经济情况及个人意愿,年轻患者要考虑到其性生活能力,其结果是医生和患者形成诊治同盟,使患者获得当前最好的治疗效果。

“两大核心”:证据要分级,推荐有级别。对所获得的证据质量要分级对待和使用,分级的目的就是要找出目前“最佳”的证据。

循证医学实践一般分成5个步骤(简称5A):

1. asking 将患者的问题按PICO原则分解成可以回答的问题。

P: 患者或患者群(patient or population),属于哪类患者或疾病?

I: 干预或暴露(intervention or exposures),即想为患者做何种检查或拟用何种治疗方法?什么因素可能影响患者的预后?

C: 对比(comparison),干预与什么相比较(是两种药物之间选择一种或与安慰剂对比或两种诊断试验选择一种)?不一定每个临床问题都需要对比。

O: 结果(outcome),希望达到什么结果(是缓解或消除症状、减少不良反应、改善功能或增加生活质量评分)?

2. accessing 查找资料回答问题。查找可获得的最好的证据信息,是本节要回答的主要问题,即在哪里查找最佳证据?如何查找最佳证据?

3. appraising 严格评读文献,对获得文献(证据)的效度与重要性进行评价。

4. applying 是否可应用到患者身上。整合严格评价的证据、医生经验、患者个体实际情况等三大层面(要素)。

5. auditing 对1~4步的执行情况进行评估。

(二) 证据级别

证据是循证医学的基石,遵循证据是循证医学的本质所在。循证医学中的证据主要指临床人体研究的证据,即以病因、诊断、预防、治疗、康复和预后等方面为基础的研究结果,尤其是以患者为中心的临床研究结果,包括一次文献和二次文献。

临床研究证据分级是循证医学所提出的要求,根据其来源、科学性和可靠性分为五级(可靠性依次降低):①大样本多中心RCT或者收集这些RCT所作的系统评价和/或荟萃分析;②单个的大样本RCT;③设有对照组的临床试验;④无对照组的系列研究;⑤专家意见、描述性研究和病案报告。

其中以高质量的原始临床研究证据和联合这些临床研究证据进行的二次研究所获得的二次研究证据(如系统评价)为质量最高的证据为一级,而基于专家或基础研究的证据级别最低为五级。这种分级对于临床医生使用证据很有帮助。

循证医学的证据也可以按加工的深度不同,将循证医学的证据资源用“5S”金字塔结构来表示(图 8-24)。

1. 研究(studies) 期刊中的原始研究,是基础,是一次文献。是对有关病因、诊断、治疗和预防等的原始研究得出的结论。主要包括通过随机对照试验、非随机对照试验、半随机对照试验、交叉试验、队列研究、横断面调查和病例对照研究等方法获得的研究结果。主要从文献数据库如 PubMed、EMBase、SinoMed 等数据库系统获取。

2. 综述(syntheses) 指系统综述。主要从 Cochrane Library、Agency for Healthcare Research and Quality (AHRQ, USA)、National Institute for Health and Clinical Excellence (NICE, UK)、Database of Abstracts of Reviews of Effects (DARE, UK) 这些数据库系统获取。

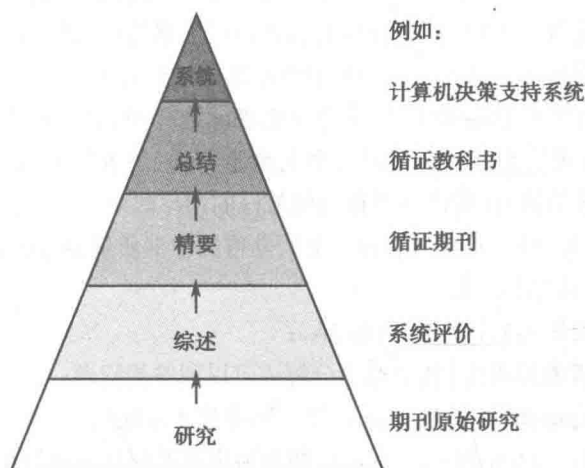


图 8-24 证据来源 5S 金字塔

3. 精要(synopses) 指出现在循证期刊对单个原始研究和系统综述的简要描述文献。从 *Evidence Based Medicine*、*Evidence Based Nursing*、*Evidence Based Paediatrics*、*Evidence based Orthopaedics and Trauma* 等期刊可以获取,代表数据库为 OVID 提供的 ACP Journal Club(美国医师协会的循证医学资料库)。

4. 总结(summaries) 临床主题的证据总结。是整合来自单个精要、系统综述或原始研究甚至它们的总和,针对某一具体疾病提供有关其治疗选择的全面证据。代表性的有 BMJ Clinical Evidence、PIER、Up to Date、Dynamend、Harrison's Practice、Essential Evidence Plus (EEP)、Cliniguide 等系统,多为证据资源综合性网站。

总结是将与某一主题各方面相关的精要、综述或研究汇总分析,产生的新证据文献,是一种临床决策辅助工具。其检索系统和功能比较完善,使用方便,可以快速解决临床问题。

精要、综述、原始研究,通常只评价治疗的一个方面(如治疗某病的某一具体药物或药物种类,如治疗高血压的血管紧张素转化酶抑制剂),由决策者自行整合,而对于原始研究,需要决策者自己去严格评价证据。

5. 系统(systems) 指能将患者个体的信息与研究证据的适用相匹配的计算机决策支持系统。系统将电子病案中的特征与当前可得的最好证据(如临床实践指南、技术评估等)自动链接,并提醒或告知医护人员治疗的关键所在。

精要、综述、总结按文献加工层次来讲,属于二次临床研究证据。主要包括系统综述、临床



实践指南、临床决策分析、临床证据手册、卫生技术评估及卫生经济学研究等。公认最好的二次临床证据数据库为 Cochrane 图书馆。

“5S”证据结构中,证据可靠性从高到低依次为:系统→总结→精要→综述→研究。因此,在循证医学实践中,检索证据用于指导临床决策时,应首先从“5S”的最高层开始,依次退而求其次。

(三) 证据资源

目前有大量可供医学研究证据查询的来源,包括原始研究数据库、二次研究数据库、期刊、指南、卫生技术评估手册等。一般可以分为以下几类:

1. 证据评价类数据库检索系统 以 Cochrane Library (简称 CL) 为代表。

(1) Cochrane Library: 由国际 Cochrane 协作网提供。CL 是临床疗效研究证据的基本来源,也是目前临床疗效研究证据的最好来源,是 Cochrane 协作网 (<http://www.cochrane.org>) 的主要产品。

(2) 循证医学评价数据库 (Evidence-Based Medicine Reviews, EBMR): 由 OVID 公司提供。包括 CL 的全部数据和美国医师协会循证资料库,并与 Medline 和 OVID 收录的杂志全文相链接。

(3) 评价与传播中心数据库 (<http://www.crd.york.ac.uk/crdweb/>): 包括了 DARE、NHSEED 和 HTA 三个数据库,由英国约克大学卫生评估与传播中心 (Centre for Reviews and Dissemination, CRD) 制作。

(4) 中国循证医学 /Cochrane 中心数据库 (<http://www.chinacochrane.org/>)。

(5) 中国卫生技术评估数据库 (http://lib-chta.sgst.cn/CN_HTA_Papers/search.asp): 包括国内和国外有关卫生技术评估的论文。

2. 原始研究类数据库 如 PubMed、EMBase、SinoMed。

3. 临床实践指南类 临床实践指南有基于专家共识的临床实践指南 (consensus based guideline) 和基于证据的临床实践指南 (evidence-based guideline)。

(1) 美国国立指南交换中心库 (National Guideline Clearinghouse, NGC): 该指南收录范围广,数量多,各种类型的指南均有,指南的质量需用户自己去评价 (<http://www.guideline.gov/>)。

(2) 基于证据的指南: 英国国家卫生和临床示范研究所 (National Institute for Health and Clinical Excellence, NICE) 临床实践指南库 (<http://www.nice.org.uk>), 苏格兰大学校际指南网络 (Scottish Intercollegiate Guidelines Network, SIGN) 临床实践指南 (<http://www.sign.ac.uk>) 等。

(3) NIH 共识发展项目数据库 (consensus development program, CDP): NIH 将共识会议法引入医学研究领域,用以辅助复杂的决策过程 (<http://consensus.nih.gov/>)。当治疗意见和诊断方法选择等存在矛盾时,共识会议法可以发挥所长。由各相关专家、群体、代表等以投票、排序或其他达成共识的互动方法,针对决策或研究发现进行评估,再由会议委员会讨论,将这些多元化的决议整合,形成“共识声明”或“推荐意见”。这些“声明”和“意见”也是很好的临床证据。

4. 证据总结类数据库 以 BMJ 临床证据在线 (Clinical Evidence, CE) 为代表。其他还有 PIER、Up to Date 等。

5. 循证医学集成搜索引擎 以 SUMsearch (<http://sumsearch.uthscsa.edu>) 和 TRIP 数据库 (<http://www.tripdatabase.com/index.html>) 为代表。

6. 循证医学专业杂志 属于精要类证据资源。

(1) 循证医学杂志 (Evidence-Based Medicine, EBM): 双月刊,由 BMJ 和美国内科医生学院 (American College of Physicians, ACP) 联合主办。为医疗卫生工作者自大量的国际性医学杂志中筛选和提供全科、外科、儿科、产科和妇科方面的研究证据。

(2) 美国内科医生学院杂志联合 (ACP Journal Club): 双月刊,由 ACP 和美国内科协会 (American Society of Internal Medicine, ASIM) 联合主办。旨在通过筛选和提供已出版的研究报

道和文献综述的详细文摘,使医疗卫生工作者掌握治疗、预防、诊断、病因、预后和卫生经济学等方面的重要进展。

(3) Bandolier: 单月刊,由 NHS 主办。主要提供干预疗效方面的最佳证据。

(4) 循证护理杂志 (*Evidence-Based Nursing*): 季刊,由英国皇家护士学院和 BMJ 联合主办。是一个提供与护理相关的最好研究和最新证据的高质量国际性新杂志。

(5) 循证卫生保健杂志 (*Evidence-Based Health Care*): 季刊,由英国出版。旨在为健康卫生管理者和决策者提供健康保健金融、组织和管理方面的最佳证据。

(6) 中国循证医学杂志。

(四) 证据源选择

循证医学证据检索的重点是各种数据库,目前能提供临床使用的数据库较多,只有针对临床需要,正确设计检索策略,并结合现有条件,才能确定什么是当前最相关的数据库和基本相关的数据库。

1. 需要用二次研究证据解决临床问题,首先选择 CL、CE、EBMR、CRDD 和 Guideline 及期刊,评价干预措施疗效首选 CL。

2. 检索临床实践指南,首选 National Guideline Clearinghouse (NGC)。

3. 需对原始资料进行二次研究,应首选 PubMed、EMBase、SinoMed 等数据库,查找原始研究证据。

4. 检索高质量原始研究文献,可选如 *Lancet*, *BMJ* 和 *JAMA* 等综合性医学杂志和临床问题所涉及的相关专科杂志。

5. 需要对在研项目进行追踪的可选择 NRR 数据库。

二、循证医学资源数据库检索

(一) Cochrane Library

Cochrane Library 通过 Wiley InterScience (<http://www.interscience.wiley.com/>) 提供服务。Cochrane Library 包含 CDSR、CENTRAL、CMR、HTA 和 NHSEED 等 5 个数据库 (<http://www.thecochranelibrary.com/view/0/index.html>)。其中, CDSR、CENTRAL、CMR 是由 Cochrane 协作网专业组制作的, HTA、NHSEED 是由英国约克大学卫生评估与传播中心 (CRD) 提供的。在 Wiley 在线图书馆资源名称为 Cochrane Reviews、Other Reviews、Trials、Methods Studies、Technology Assessments、Economic Evaluations、Cochrane Groups 等。

(1) Cochrane Reviews: 即 CDSR, Cochrane 系统评价数据库 (Cochrane Database of Systematic Reviews, Cochrane Reviews)。收录由 Cochrane 协作网 50 余个系统综述专业组在统一工作手册指导下完成的系统综述,包括系统综述全文 (Review) 和研究方案 (Protocol),并随着读者的建议和评论以及新的临床试验的出现不断补充和更新。

(2) Trials: 即 CENTRAL, Cochrane 中心临床对照试验注册数据库 (Cochrane Central Register of Controlled Trials, Clinical Trials)。资料来源于协作网各系统评价小组和其他组织的专业临床试验资料库以及在 MEDLINE 上被检索出的随机对照试验 (RCT) 和临床对照试验 (CCT)。还包括了全世界 Cochrane 协作网成员从有关医学杂志会议论文集和其他来源中收集到的 CCT 报告。

(3) Methods Studies: 即 CMR, Cochrane 方法学注册库 (Cochrane Methodology Register)。主要收录有关对照试验方法和系统评价方法学的相关文献的书目信息。信息来源包括期刊文献、图书和会议录等 (这些文献主要来源于 MEDLINE 数据库和人工查找所得)。

(4) Technology Assessments: 即 HTA 数据库,卫生技术评估数据库 (Health Technology Assessment Database)。收集来自国际卫生技术评估协会网 (INHTA) 和其他卫生技术评估机构



提供的完成和正在进行的卫生技术评估。

(5) **Economic Evaluations**: 即 NHSEED, NHS 经济评估数据库 (NHS Economic Evaluation Database)。有关成本效益、成本效能的分析。可协助决策者从全世界搜集系统性的经济性评估,并鉴定其质量及优缺点。

Cochrane 图书馆资源的文献和记录状态有多种类型: ① **Review**: 有完整的结果和讨论、数据分析和图表的系统综述; ② **Protocol**: 系统综述的准备大纲(方案), 包括背景、原理说明和方法; ③ **Methodology**: 系统综述方法学研究的全文文献; ④ **Comment**: 含有评论或批评的完全系统综述。读者可以提出自己的评论, 这些评论会被整合到作者的回答与回馈中; ⑤ **New**: 最新一期版本中所收录的最新的 protocol 或 review; ⑥ **New search**: 对已经出版的系统综述有关研究的新的检索; ⑦ **Conclusions changed**: 对最新一期版本中系统综述的结论有重要的改变; ⑧ **Major change**: 重大变化, 对最新一期版本中的系统综述进行了修订影响到了其内容的改变; ⑨ **Withdrawn**: 被撤销的 review 或 protocol, 通常是因为缺乏相关活动或更新, 撤销的理由会在该篇文章上详细说明。这些类型在其结果显示时以图标形式提示(图 8-25), 在高级检索的记录状态限定选项 (Restrict Search by Record Status) 也提供选择。

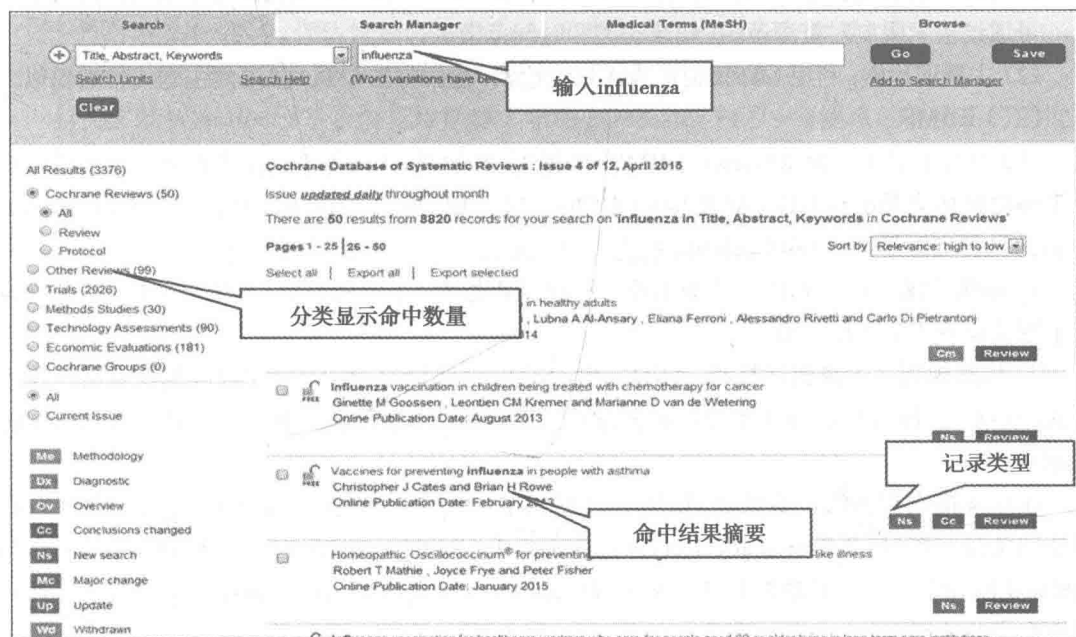


图 8-25 Cochrane 图书馆检索结果

2. **浏览** 包括从主题、新的系统综述 (New Reviews)、更新综述 (Updated Reviews)、系统评价组 (by Review Group)、临床试验 (Trials)、方法研究 (Methods Studies)、技术评估 (Technology Assessments)、经济评价 (Economic Evaluations) 等资源类型为入口 (图 8-26)。

3. **检索** 包括普通检索和主题词途径检索等。

检索规则: ①逻辑运算符: AND、OR、NOT; ②截词符: *; ③位置符: NEAR/ 数字 (表示左右相隔多少个单词), NEXT (表示左右相邻); ④字段限制表达方式: 提问: 字段缩写 (如 H1N1: ti, 表示检索在标题中含有 H1N1 的文献); ⑤短语检索: 将短语用英文半角双引引起来, 如 "H1N1 flu"。

以上这些检索规则, 除字段限制检索只能在“检索史”状态下使用外, 均可在基本检索和高级检索模式使用。

(1) **普通检索**: 将检索词键入检索提问框即可。可点选下拉按钮设定是从题名、摘要或关

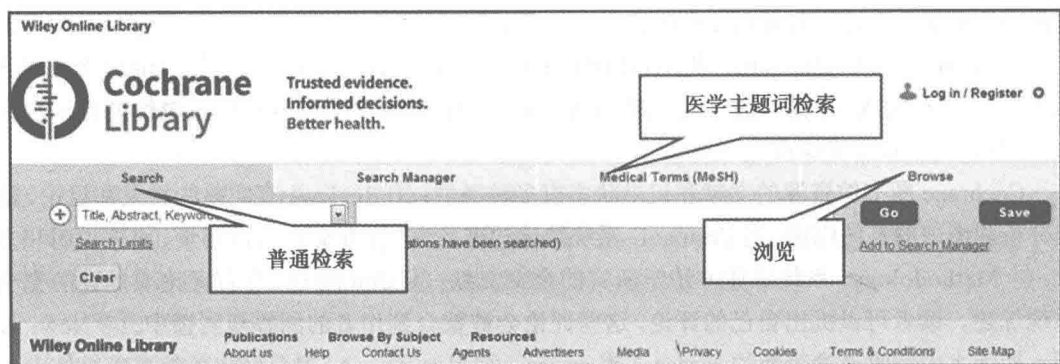


图 8-26 Cochrane 图书馆资源检索页面

关键词等字段进行检索。如在检索框中输入“influenza”，点击按钮“Go”，系统返回结果如图 8-25。返回结果包括 Cochrane 综述、其他综述、临床试验、方法学研究、卫生技术评估、卫生经济评估、Cochrane 小组等几个方面的文献，“[]”里面的数字表示命中的文献数。

可通过单击普通检索输入框前面的“+”，增加多个输入选项框，填表构建检索式进行检索。

可通过检索限制对资源范围、记录状态和时间范围进行限制。

(2) 主题词检索：利用 MeSH 词汇为入口词检索，检索方法与 PubMed 的主题词检索相似。

(二) EBMR

循证医学评价 (EBM Reviews, EBMR) 数据库通过 OVID 服务平台提供网络检索服务。收录美国医师协会循证资料库 (ACP Journal Club, 包含 Evidence Based Medicine 和 ACP Journal Club 两种循证医学期刊) 和 Cochrane 图书馆的全部资源。

1. 检索方法 同 OVID 全文数据库。提供基本检索、主题词检索、字段限制检索、高级检索和题录检索等多种检索模式。

2. 检索规则 ①逻辑运算符：AND、OR、NOT；②截词符：*、\$；③字段限制检索语法表达方式：提问. 字段缩写。(在基本检索状态，输入 hepatitis.ti., 表示检索在标题中含有 hepatitis 的文献)。

在基本检索和高级检索状态，EBMR 还提供了限定选项供选择。这些限定选项包括系统综述的资源类型 (Full Systematic Reviews、Protocols、Review Articles、New Reviews、Recently Updated Review)、临床资源类型 (Etiology、Diagnosis、Therapeutics、Prognosis)、对象等。检索时可根据需要选择。

3. 结果输出 完成检索后，在每条记录的右边有“EBM Topic Review”链接，点击该链接可打开该记录的全文，点击“Internet Resources”可进一步在 ACR (Cross Archive Searching, <http://arc.cs.odu.edu/>)、Intute (<http://www.intute.ac.uk/>)、SCIRUS (<http://www.scirus.com/>) 检索与该主题相关的循证医学资源。

(三) BMJ 临床证据库

属于证据总结类数据库，汇总来自系统评价、随机对照试验，以及经过同行评议过的观察性研究 (如无高质量证据的情况下) 的当前可得的最佳证据。BMJ 的临床证据数据库，通过 BMJ 网站 (<http://clinicalevidence.bmj.com/x/index.html>) 提供服务。

1. 浏览 CE 提供按条件状态 (conditions) 浏览检索。点击屏幕右上方的 Show Conditions 按钮，系统弹出分类标签，每个分类按字母顺序排列的主题，选择需要的项目便可打开系统综述或其他相关 (Covered elsewhere) 资源 (图 8-27)。

2. 检索 CE 只提供简单的类似搜索引擎的检索，检索结果提供一小段内容及突出显示搜索所用的关键词，显示内容概要，并按相关性排序。





图 8-27 CE 主页

3. 结果 CE 提供的证据资源(系统综述), 以标签子菜单的形式展示系统综述的详细内容, 包括干预措施效果、要点、评估分级 (GRADE table)、背景 (Background)、参考文献、学分 (Credits)、相关文献、患者信息、指南、BMJ 相关内容 (Related BMJ content) 等。

相关文献 (Citations) 是由麦克马斯特大学的 McMaster PLUS 监测系统鉴别后的证据文献 (<http://hiru.mcmaster.ca/more/>), 分为治疗与预防 (Therapy/Prevention)、质量改善 (Quality Improvement)、诊断、预后、病因学 (Etiology)、临床预测指南 (Clinical Prediction Guide) 等 6 类, 每篇证据文献由医生或护士根据相关性和新闻性给予了 1~7 颗星级评价 (图 8-28)。



图 8-28 CE 结果详细显示

(四) 临床实践指南检索

因特网上有很多临床实践指南站点, 其中美国国立指南交换中心 (NGC) 收录的指南较多, 检索功能较强 (<http://www.guideline.gov/>)。

NGC 提供浏览、基本检索、高级检索、指南比较等服务。可从 Disease/Condition、Treatment/Intervention、Measures、Guideline Index、Organization、Guidelines In Progress、Guideline Archive

等入口浏览相应的指南内容。在 NGC 的主页左上方的检索提问框内输入检索词即可完成基本检索(图 8-29)。点击 Detailed Search 链接,便可进行细节检索(也称限制检索),系统提供的多个限制选项,可使检索结果更加精确。

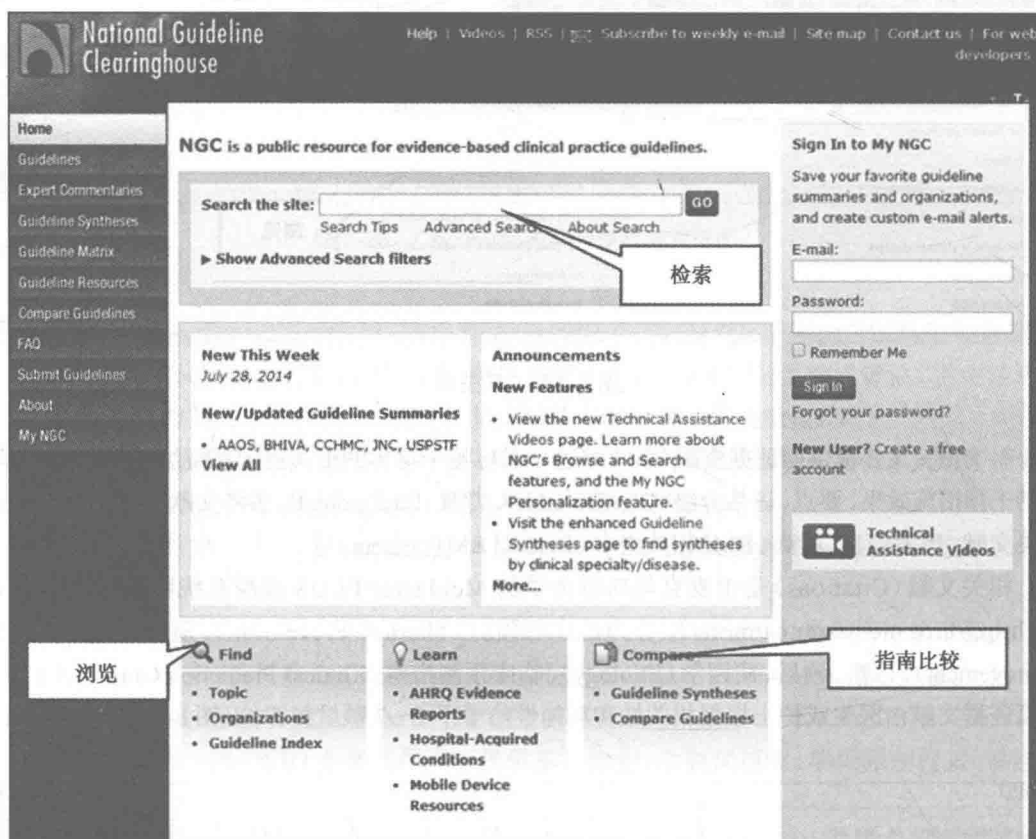


图 8-29 NGC 主页

指南比较(Compare)是 NGC 的特色功能,用户通过检索或浏览,可以任选两个或两个以上的指南进行比较。操作步骤如下:①点击检出结果指南前面的方框,加上标记;②点击“Add to My Collection”按钮,将标记的指南添加到我的收藏中去;③点击“Compare Checked Guidelines”按钮。系统自动对选择的指南从标题、时间、具体内容等方面进行列表比较。

(五) 原始研究证据检索类数据库

外文文献以 PubMed 为代表,中文以 SinoMed 为代表。

1. PubMed 可检索原始研究证据,也可检索二次研究证据。

(1) 利用临床咨询模块:临床咨询检索采用过滤器(filter)模式将无效或不适用信息滤掉,只保留有助于解决临床问题的信息。该模式不需复杂的检索策略,就可直接检索到所需的临床研究文献。

1) 系统综述:包括 PubMed 所收录的期刊上按 Meta 分析方法进行研究的系统评价文献和 Cochrane 协作网提供的部分系统综述的文摘。

2) 临床研究类文献:在检索框中输入某一疾病名称(如 hepatitis),可进一步限定在病因、诊断、治疗、预后和临床预测指南等更细的领域和选择检索范围(Scope)确定查全率(Broad, sensitive search)或查准率(Narrow, specific search),对文献进行筛选(图 8-30)。

(2) 利用结果筛选中“Article types”选项:限制文献类型为 Clinical Trial、Meta-Analysis、Practice Guideline、Randomized Controlled Trial。

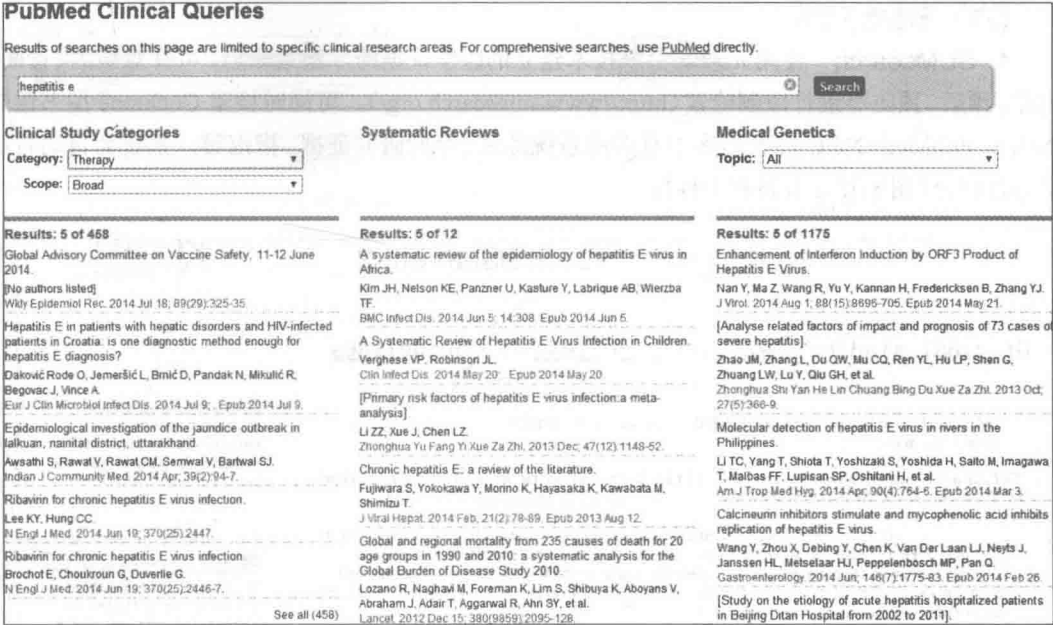


图 8-30 PubMed 临床咨询文献检索界面

(3) 直接以 systematic review、Cochrane、meta-analysis、multicenter-study 等为关键词进行检索。

2. EMBASE 在 EMBASE 的检索限制(RECORD LIMITS)中有专门的循证医学文献类型限制选项(图 8-31)。

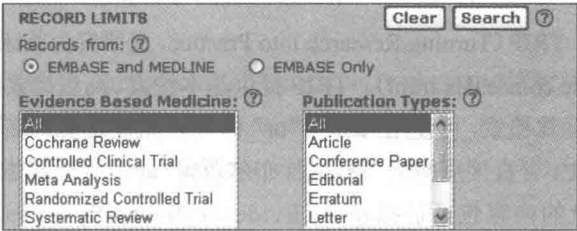


图 8-31 EMBASE 的循证医学文献检索限制选项

3. 中国生物医学文献数据库 SinoMed 是国内常用的一个综合型的生物医学文献数据库，其使用方法与 PubMed 类似。使用“循证医学”、“系统综述”、“系统评价”、“荟萃分析”、“meta 分析”、“汇后分析”等作为检索词可检索到二次证据文献；使用一些特定的术语作为检索词，并通过逻辑组配检索可获取原始研究证据文献。

(1) 与“病因学”相关的检索词或策略：随机对照试验，队列研究，横断面研究，多因素分析 [MH]，前瞻，危险因素 [MH]，风险，病例对照，健康对照，正常对照，相关，关系，有关，无关，人类 [CT]，动物 [CT]。

(2) 与“诊断”相关的检索词或策略：敏感度(性)，灵敏度(性)；特异性(度)，诊断 [SH]，诊断应用 [SH]，阳(阴)性预测值。

(3) 与“治疗”相关的检索词或策略：药物疗法 [SH]，治疗应用 [SH]，随机对照试验 [MH]，随机分配 [MH]，双盲法 [MH]，单盲法 [MH]，安慰剂，随机，盲法，对照，对比研究 [CT]，人类 [CT]，动物 [CT]。

(4) 与“预后”相关的检索词或策略：回归分析 [MH]，分层，发生率，死亡率，生存率，存活率，预后 [SH]，随访研究 [SH]，随诊，队列研究。



(六) 证据搜索引擎

1. SUMsearch 使用元搜索引擎技术建立的医学证据跨库检索平台, 可针对病因、诊断、治疗、预后、预防等进行限制检索 (<http://www.sumsearch.org/>)。可同时检索 Cochrane 图书馆的 DARE、PubMed、NGC 等数据库中有关的系统综述、原始研究证据、指南等。系统会自动将结果进行整合(图 8-32), 并按相关性排序。

Original studies Systematic reviews Guidelines

988 possible original studies PubMed found after 6 searches. The first 50 citations are:

1. Sofosbuvir and ribavirin for hepatitis C in patients with HIV coinfection. JAMA 2014 Jul 23-30;312:4. PMID: 25038354, doi: 10.1001/jama.2014.7734. Cite

Conclusion: In this open-label, nonrandomized, uncontrolled study, patients with HIV who were coinfectd with HCV genotype 1, 2, or 3 who received the oral, interferon-free combination of sofosbuvir and ribavirin for 12 or 24 weeks had high rates of SVR12. Further studies of this oral regimen in diverse populations of coinfectd patients are warranted.

Impact/quality: Accompanied by editorial; *

2. Simeprevir with pegylated interferon alfa 2a plus ribavirin in treatment-naive patients with chronic hepatitis C virus genotype 1 infection (QUEST-1): a phase 3, randomised, double-blind, placebo-controlled trial. Lancet 2014;:. PMID: 24907225, doi: 10.1016/S0140-6736(14)60494-3. Cite

Conclusion: Simeprevir once daily with peginterferon alfa 2a and ribavirin shortens therapy in treatment-naive patients with HCV genotype 1 infection without worsening the adverse event profiles associated with peginterferon alfa 2a plus ribavirin.

Show evidence search details (access to earlier iterations If results displayed are too restrictive)

You found 50 or more articles. Consider performing your search at Trip Database. Trip curates the articles it indexes so you may get a smaller number of more helpful documents.

Wednesday, 30 Jul 2014 06:58:42 at SUMSearch

图 8-32 SUMsearch 检索结果总览

2. TRIP 数据库 TRIP (Turning Research into Practice, 变研究为临床数据库), 始于 1997 年 (<http://www.tripdatabase.com/index.html>)。TRIP 提供基本检索、高级检索和 PICO 检索。基本检索状态下可进行逻辑运算检索, 可使用“and”、“or”和“()”组配运算。在高级检索状态, 系统按检索顺序显示检索历史, 可直接用检索序号进行布尔逻辑“and”运算检索。

TRIP 数据库收录的内容包括证据精要 (Evidence-Based Synopses), 临床问题 (Clinical Q&A), 系统综述, 北美、欧洲、其他指南, 对照试验 (Controlled trials), 重要的原始研究 (Key Primary Research), 扩展的原始研究 (Extended Primary Research) 病例报告 (Case reports)、电子书籍等方面的论文(图 8-33)。

TRIP 和 SUMsearch 可同时检索多种资源, 并根据证据来源组织检索结果。在不熟悉针对某一具体临床问题采用哪种循证资源最好, 或拥有的资源无法解决手头的问题时, 可借助 SUMsearch 和 TRIP“联合搜索引擎”来检索最佳证据。

3. Doctors Desk/PCEL 由伦敦圣乔治医院医学院社区卫生系创建。可检索循证医学方面的指南、系统评价或研究论文。指南来自 SIGN, NICE, PRODIGY 和 NeLH, 循证医学文章来自 Clinical Evidence, Bandolier, Cochrane Review, ACP Journal Club 和 Effectiveness Matters, 检索功能较为完善 (<http://drsdsk.sghms.ac.uk>, <http://www.pcel.info/index.php?fuse=search.Ebmprimer>)。

(七) 常用医学统计数据网络资源

1. WHO 统计信息系统 <http://www.who.int/whosis/zh/index.html>。世界卫生组织流行病学数据系统 (WHOSIS) 提供其统计学和流行病学数据信息。系统可利用关键词途径检索 WHO 网站上的统计数据, 也可以通过该网页链接到各专门数据库。

可查询死亡原因、婴幼儿死亡原因、平均寿命、年龄标准化死亡率、世界各国流行病学信息 (包括疾病暴发、出现的重要传染病等情况) 以及全球监测与控制疾病的有效措施等信息。





图 8-33 TRIP 检索结果分类显示

2. TDRdata <http://www.tdrdata.com/default.aspx>, 全球流行病学数据库和药学研究。包括美国住院、急症、医院、医生、危险因素调查数据 (BRFSS) 以及欧盟 25 个国家住院患者费用负担率。

3. 卫生部卫生统计数据 <http://www.nhfpc.gov.cn/zwgkzt/pwstj/list.shtml>

4. 中国 CDC 卫生统计 <http://www.chinacdc.cn/tjsj/>

5. 香港食物及卫生局统计数据 <http://www.fhb.gov.hk/statistics/chs/statistics.htm>

(赵文龙)

第五节 药学信息检索

药学是由化学、医学、生物学等相关学科知识综合而成的一门交叉学科, 药学信息则涵盖了药学各个学科、专业领域的内容, 包括药品的研制、生产、流通、使用和管理, 横跨科研、工业、商业、卫生、教育、管理等行业和系统。药学信息检索从资源类型的角度来看包括药学文献、专利、产品等信息检索, 最重要的是药学文献信息的检索; 从专业角度来看, 主要包括药物制备、药理、毒理、药物流行病学 (不良反应)、临床应用等。

一、综合性药学文献数据库

(一) 中国中医药数据库检索系统

中国中医药数据库检索系统 (<http://cowork.cintcm.com/engine/windex.jsp>) 由中国中医科学院中医药信息研究所建设, 包括中医药期刊文献数据库、疾病诊疗数据库、各类中药数据库、方剂数据库、民族医药数据库、药品企业数据库、各类国家标准数据库 (中医证候治则疾病、药物、方剂) 等相关数据库 (图 8-34)。所有的数据库都可联网使用, 部分数据提供英文版。

中医药数据库检索系统可以实现单库检索与跨库检索。

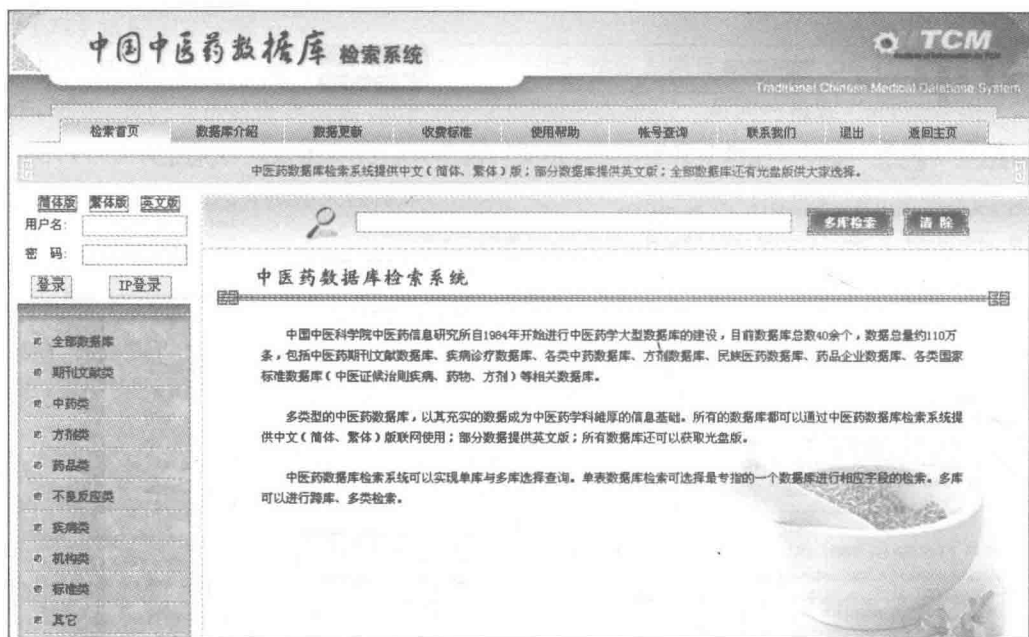


图 8-34 中国中医药数据库检索系统

(二) EMBASE 的药物信息检索

药物信息是 EMBASE (<http://www.embase.com>) 最具有特色的资源, 其药学文献标引、检索独到, 与药物相关的有 drugs & pharmacology 及 pharmacy 两个子数据库。前者收录了 1990 年以来关于药物或研制中的药物的临床、药理、药代动力学、药效、副作用等内容, 后者主要收录 1988 年以来的有关药物生产、药物赋形剂、药物释放系统等内容的文献。涵盖药物研究、药理学、制药学、药剂学、毒物学、人体医学(临床和实验)等方面, 还特别提供药物及医疗器械的公司和制造商的系统索引。

EMBASE 提供的“药物检索”是检索药物文献的主要途径(图 8-35)。可通过化学名称、商品名和制造商名称检索。建议从 Emtree 入口以药物名称为检索词检索。

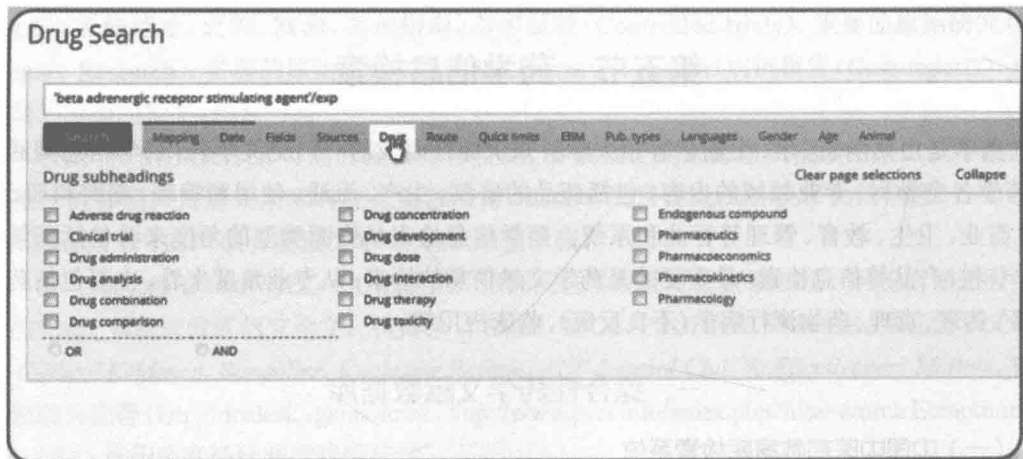


图 8-35 EMBASE 的药物检索主界面

药物检索提供自动转换到优选术语、药物作为关键词检索、扩展检索和以某药物为研究重点的专门检索; 提供药物副主题, 如药物副作用反应、临床试用、药物分析等; 提供常用给药方式(route)的检索, 如口服、肌内注射、静脉注射等。



(三) 国际药学文摘

国际药学文摘(International Pharmaceutical Abstracts, IPA)由美国医院药师学会(American Society of Health-System Pharmacists, ASHP)提供。收录了1970年以来世界范围的850多种主要期刊和美国国内的所有药剂期刊(包括大部分化妆品出版物)。1985年起收录范围扩大到有关论述相关条例、法律,人力资源和薪资方面报道的药物期刊。1988年,IPA开始收录ASHP的主要会议论文的文摘,现在还收集美国药学会(American Pharmaceutical Association, APhA)和美国药学院协会(American Association of Colleges of Pharmacy, AACP)年会推荐的论文文摘。该文摘提供包括研究设计、患者数量、剂量、剂型、剂量时间表等临床研究报告的信息。

在线DIALOG系统可联机试用。CAS、OVID等都提供在线检索服务。OVID系统的IPA的检索包括基本检索、字段检索、索引检索等。

IPA中有关药物的分类是依据美国医院药典服务(American Hospital Formulary Service, AHFS)制定的药物学/治疗学分类系统进行分类的。每个文献记录都列出了药物的类别及其分类号。在OVID系统可通过“药物/治疗分类”(Pharmacologic/Therapeutic Classification, PC)字段完成。药物分类系统的分类号和药物类别都是PC的有效检索内容,PC是非限制字段,可以用相应的药物分类号或分类类目检索。格式为:“分类号.pc.”或“药物类别.pc.”。

(四) 毒理学网络

毒理学网络(Toxnet, <http://toxnet.nlm.nih.gov>)提供了包括毒理学在线(TOXLINe)、有害物质数据库(HSDB)、化合物别名和结构数据库(ChemIDplus)、风险信息集成系统(IRIS)、基因突变毒理学(GENETOX)、化学致癌研究信息系统(CRIS)、发育和生殖毒理学(DART)、化学物环境毒理学评价(TRI)等数据库。

TOXLINe和HSDB是NLM使用频率最高的两个数据库。TOXLINe是有关药物和化学物质的毒理学、药理学、生理学等内容的文献型数据库,是常用的药学文献数据库之一。HSDB是关于化学物质对人类和动物的毒性、安全、使用以及评价的文献专题数据库。

联机检索TOXNET的数据库比较简单,可以选择某一个或全部数据库进行检索(图8-36)。在检索提问框中输入关键词,点击Search按钮,系统就会将命中结果显示在屏幕,用户可以进

United States National Library of Medicine

TOXNET

Toxicology Data Network

TOXNET PDA Access SIS Home About Us Site Map & Search Contact Us

► Env. Health & Toxicology ► TOXNET

TOXNET - Databases on toxicology, hazardous chemicals, environmental health, and toxic releases.

Select Database

- ChemIDplus ☐
- HSDB ☐
- TOXLINe ☐
- CCRIS ☐
- DART ☐
- GENETOX ☐
- IRIS ☐
- ITER ☐
- LactMed ☐
- Multi-Database ☐
- TRI ☐
- Haz-Map ☐
- Household Products ☐
- TOXMAP ☐

Search All Databases

Enter term(s) to search all databases.

(e.g. asthma air pollution, ibuprofen fever, vinyl chloride)

TOXNET Search Options

- Search all databases: Enter term(s) in box above
- Search a specific database: Click database at left
- Database description: Click on the ☐

Env. Health & Toxicology

Portal to environmental health and toxicology resources

Support Pages

- Help
- TOXNET FAQ
- TOXNET Update Status
- Fact Sheet
- Database Description
- Training Manual & Schedule
- News

图 8-36 TOXNET 主页

一步处理检索结果(标记、细览、打印、下载等)。

(五) 美国化学文摘

美国化学文摘(CA)从化学的角度提供与药物相关的文献,包括药物合成、筛选、分析、制备、分析、药理、毒理等。

二、与临床处方相关的药物信息

(一) 《默克手册》与《默克索引》

《默克手册》(Merck Manuals)是系列药物工具书,包括《默克诊疗手册》《默克家庭诊疗手册》《默克老年手册》《默克索引》等。最著名的是《默克诊疗手册》,是很有价值的参考书籍,中文版由默沙东中国公司提供免费在线阅读(<http://www.msdchina.com.cn/manual/manual.html>),但《默克索引》不提供免费阅读。

《默克索引》(Merck Index),已出版 14 版,是查找化学品、药物以及生物化学品信息的首选。是世界著名的有关化学物质、药品及生物制品的百科全书。每一条记录评述一种单一存在的化学物质或一组非常密切相关的化合物。记录的内容包括:分子式和分子量、标准化学名称(包括 CAS 采用的名称)、普通名称和俗名、商标及其拥有者、公司代码、CAS 登记号、物理和毒理数据、治疗应用、商业应用等(图 8-37 是其光盘版的主界面及检索结果)。

The screenshot displays the 'The Merck Index' software interface. On the left is a sidebar with navigation options: 'Compound Search', 'Name Reactions', 'Additional Tables', 'Read Me First', 'Technical Support', and 'Documentation'. The main window shows search results for a compound. At the top right is a chemical structure diagram. Below it, the text reads: 'THE MERCK INDEX is a trademark of Merck & Company Incorporated, Whitehouse Station, New Jersey, USA and is registered in the United States Patent and Trademark Office. Published originally as Merck Index, Thirteenth Edition. Copyright 2001 by Merck & Co., Inc., Whitehouse Station, New Jersey, USA. All Rights Reserved.' The search results section includes: 'Monograph Number: 8063', 'Title: Tolimide', 'CAS Registry Number: 135755-83-2', 'Molecular Formula: C₁₂H₁₄N₂O₂', 'Molecular Weight: 226.26', 'Percent Composition: C 65.50%, H 6.37%, N 17.71%, O 10.40%', 'Infrared (KBr): 1650 (C=O), 1550 (C=N), 1450 (C=C), 1350 (C-O), 1250 (C-N), 1150 (C-H), 1050 (C-H), 1000 (C-H), 950 (C-H), 900 (C-H), 850 (C-H), 800 (C-H), 750 (C-H), 700 (C-H), 650 (C-H), 600 (C-H), 550 (C-H), 500 (C-H), 450 (C-H), 400 (C-H), 350 (C-H), 300 (C-H), 250 (C-H), 200 (C-H), 150 (C-H), 100 (C-H), 50 (C-H), 0 (C-H)'. The bottom right corner shows 'Therap. Cat.: In treatment of male erectile dysfunction'.

图 8-37 《默克索引》及其记录格式

(二) 医师案头参考

医师案头参考(PDR, <http://www.pdr.net/>)由美国 Thomson 公司建立的供美国医师、护士和大众查询药物信息的网络。对于医师、护士、医学生要求必须是在美国本土的才可使用(注册时要求提供相应的 DEA 码或学校名称)。PDR 同时也提供 PDRhealth (<http://www.pdrhealth.com/>)免费版, PDRhealth 设有 Drugs & Supplements、Disease&Conditions、Clinical Trials 和 Surgery 等栏目。

Drugs & Supplements, 查找 FDA 批准的处方药、非处方药和植物药的信息,包括药物商品名、药物相互作用、副作用、处方注意事项等。可以按药名索引浏览检索,也可以直接输入药名检索(图 8-38)。

The screenshot shows the 'Prescription Drugs A-Z' search interface on the PDRhealth website. At the top, there are tabs for 'Prescription Drugs', 'Over the Counter (OTC)', and 'Herbals & Supplements'. Below the tabs is a search bar with the text 'Prescription Drugs A-Z' and a 'Search' button. A grid of letters (A-Z) is displayed, with 'A' through 'Z' in individual boxes. Below the grid, there is a pagination bar showing '1 - 25 of 77' and buttons for 'Next' and 'Last'. At the bottom, there are three input fields with the text 'Abilify', 'Accolate', and 'Accupril'.

图 8-38 PDRhealth 药物检索界面

(三) 网上处方药物索引

网上处方药物索引(Rxlist, <http://www.rxlist.com>)可通过药品名称(Drugs A~Z)、颜色、外形、标记号(imprint)、商标等药丸识别信息(Pill identifier), 疾病或症状(Diseases & Conditions)入手查找系统收录的 5000 余种药物的相关信息。该系统提供的检索结果十分详细, 包括处方组成、原辅料所符合的标准、药物作用、作用原理(简述)、结构式、分子式、分子量、性状、溶解情况以及该药物剂型、剂量。此外还包括临床药理、适应证、剂量和用法、包装、警告、禁忌证、注意事项、副作用、药物相互作用、过量、患者信息等内容。

(四) 美国药物不良反应监测网

美国药物不良反应监测网(MedWatch, <http://www.fda.gov/Safety/MedWatch/default.htm>)由 FDA 建立, 用于药品上市后的质量跟踪和监察, 以便及时发现临床试验中尚未揭示的不良反应。

该网站提供医药产品安全信息(Medical Product Safety Information)查询, 可以查看药物、医疗器械的安全信息。还可通过直报系统(MedWatch Online Voluntary Submission Form 3500)向 FDA 提交药物的不良反应报告。

(五) 药品查询

1. CDER 的 drug@FDA 可获取美国 FDA 批准的上市药品的说明书信息(label information), 输入通用名或是商品名或日期, 点击后就进入药品信息的网页, 可查看药品的批准历史、治疗等效性的信息。

2. DailyMed FDA 批准的在售处方药说明书(<http://dailymed.nlm.nih.gov/dailymed/about.cfm>)。

3. eMC 英国和欧盟许可的药品信息(<http://emc.medicines.org.uk/>), 面向医生或护士提供专业信息(SPC), 面向患者提供简单易懂的药品说明书信息(PIL)。还可通过药物的活性成分(ACTIVE INGREDIANTS)进行检索(图 8-39)。

HOME

MEDICINES

COMPANIES

LATEST UPDATES

ABOUT EMC

SIGN UP

LOG IN

Enter medicine name or company

GO

Advanced search >

Search results for medicines beginning with A

812 results found | View 200

Discontinued

Black Triangle

Show history

FILTER BY LETTER	MEDICINE NAME	ACTIVE INGREDIENTS	COMPANY NAME
A B C	Abidec Multivitamin Drops Updated 18-Jan-2013	ascorbic acid, ergocalciferol, nicotinamide, pyridoxine, riboflavin, thiamine, vitamin a	Omega Pharma Ltd
D E F	Abidec Multivitamin Drops Updated 21-Jan-2013	vitamin a - palmitate, vitamin b1 - thiamine mononitrate, vitamin b2 (riboflavine), vitamin b6 (pyridoxine hydrochloride)	Omega Pharma Ltd
G H I	Ability 7.5 mg/ml solution for injection (intramuscular) Updated 03-Jan-2014	aripiprazole	Otsuka Pharmaceuticals (UK) Ltd
J K L	Ability 7.5 mg/ml solution for injection (intramuscular) Updated 03-Jan-2014	aripiprazole	Otsuka Pharmaceuticals (UK) Ltd
M N O	Ability Maintena 400 mg powder and solvent for prolonged-release suspension for injection Updated 15-Apr-2014	aripiprazole	Otsuka Pharmaceuticals (UK) Ltd
P Q R	Ability Maintena 400 mg powder and solvent for prolonged-release suspension for injection Updated 15-Apr-2014	aripiprazole	Otsuka Pharmaceuticals (UK) Ltd
S T U			
V W X			
Y Z 0-9			

图 8-39 eMC 药品浏览检索页面

(一) PubChem

PubChem 数据库 (<http://pubchem.ncbi.nlm.nih.gov/>) 提供小分子生物活性的相关信息, 是 NIH 的分子文库计划 (MLI) 的重要组成部分, 是 MLI 的资料存储库。建立 PubChem 的目的在于促进公共部门对小分子数据库的使用, 拓展小分子技术的研究, 以便更深入广泛地理解与健康 and 疾病相关的基因、细胞及有机体的功能, 帮助生物医学研究人员发现具有疾病治疗潜在活性的分子结构。PubChem 包括 BioAssay (化学成分的生物活性筛选资料)、Compound (小分子化学物质结构)、BioActivities (具有生物活性的化学物质筛选数据)、Target (药物基因靶向)、Patents (药物专利) 五个数据库, 通过 NLM 的 Entrez 平台提供服务, 检索规则与方法同 PubMed。

PubChem 提供的检索服务与 SciFinder 或 CAS Registry (<http://www.cas.org/EO/regsys.html>) 提供的服务相似, 可以通过多种查询条件搜索超过一千四百多万种化合物 (Thomson Pharma 的二百多万个化学结构也纳入了 PubChem 数据库中) 的相关信息, 并可通过 FTP 免费下载 PubChem 整个数据库 (<ftp://ftp.ncbi.nlm.nih.gov/pubchem/>)。

PubChem 的化合物检索还提供结构检索和分子式检索。

1. 结构检索的入口 提供相似度 (Similarity)、亚结构 (Substructure)、母体结构 (Superstructure) 检索。

(1) 相似度: 输入化学物质的部分或完整结构检索药物。相似度指化学物质的结构相似的百分率, 检索时可指定与某结构的相似度。

(2) 亚结构、母体结构: 与相似度相似。

2. 分子式检索 输入分子式检索, 如 C₉H₈O₄。

3. 结构式辅助输入 相似度检索、亚结构和母体结构检索,都需要输入结构式,结构式是图形模式,与文本输入不同。PubChem 提供了结构式辅助编辑工具,具体使用方法如下:

(1) 点击检索框前面的图标,系统弹出结构式编辑框(图 8-40)。

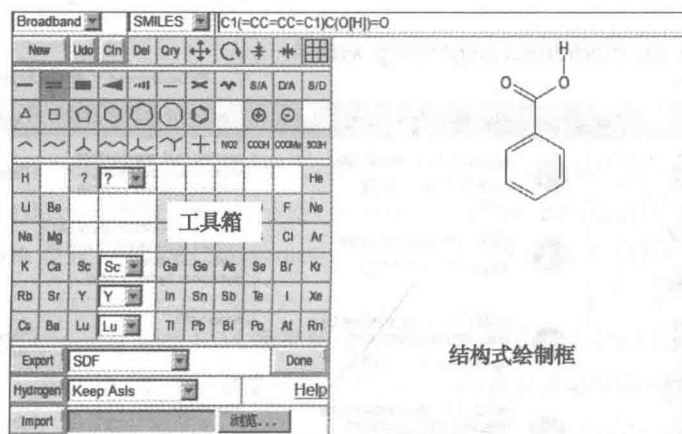


图 8-40 PubChem 结构式绘制对话框

(2) 从工具箱中选定需要的元素。

(3) 用鼠标将选定的元素拖曳到结构式绘制框适当的位置。

(4) 绘制结构式的同时,系统实时返回结果(图 8-41)。

4. 检索结果 PubChem 详细结果显示采用标签导航模式(图 8-42), 点击左栏的标签, 自动定位显示相关内容。点击 可以不同的格式下载化学物的结构。

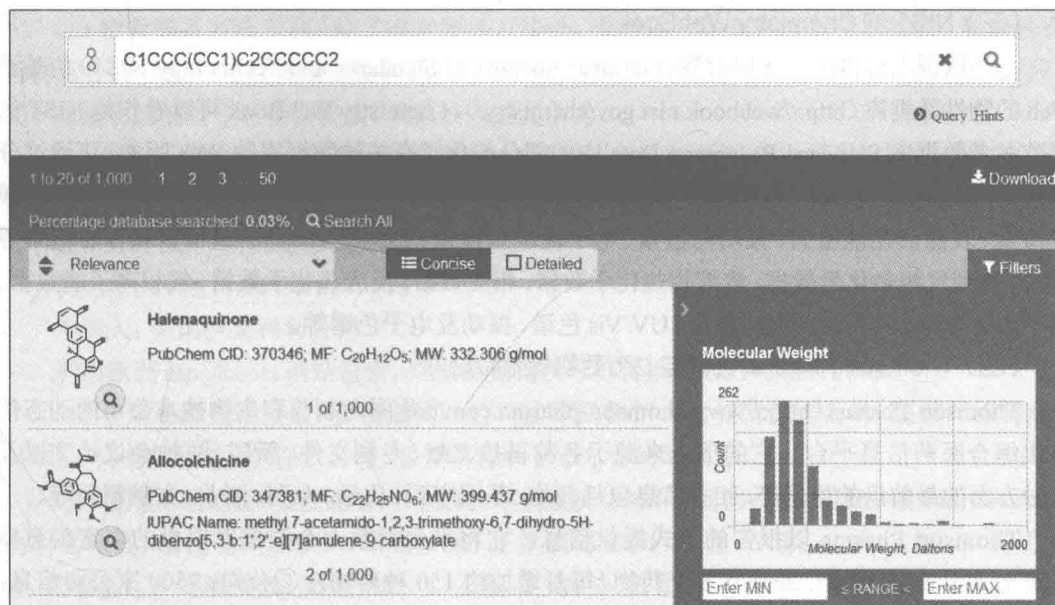


图 8-41 PubChem 检索结果

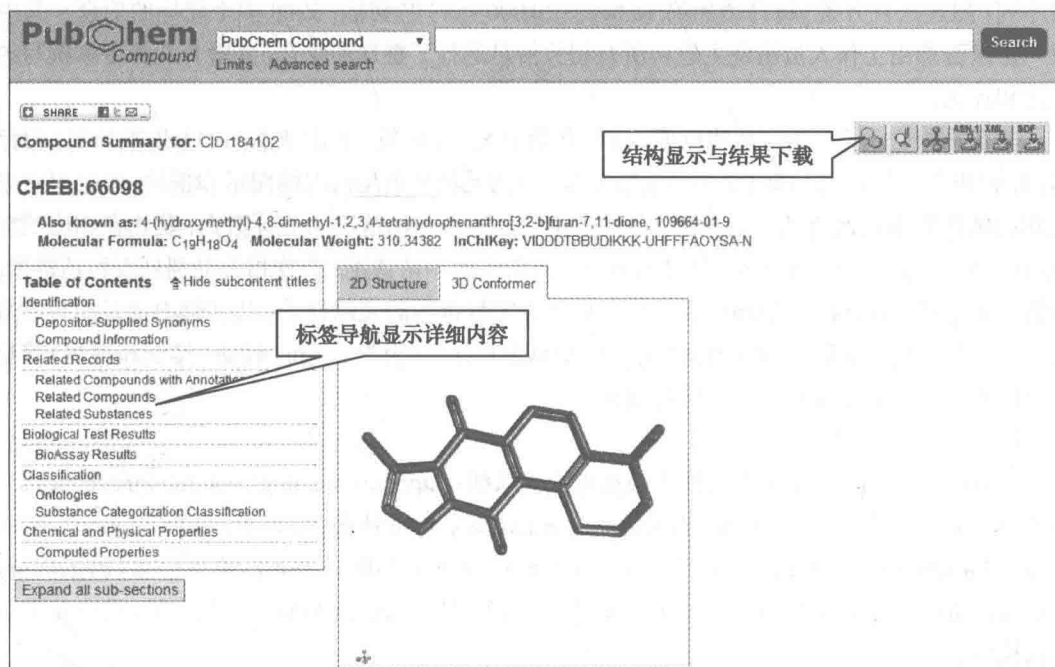


图 8-42 PubChem 结果详细显示

(二) ClinicalTrials.gov 临床药物试验研究信息

由美国国立卫生研究院 (NIH) 与所有的 NIH 研究所和 FDA 合作建立 (<http://clinicaltrials.gov/>)。ClinicalTrials.gov 于 2000 年 2 月开通, 目前美国 50 个州和 187 个国家 172 445 项研究 (截至 2014 年 8 月 6 日)。给患者、患者家属、卫生保健专业人员及公众提供有关在各种疾病中进行的药物临床研究信息。

从 ClinicalTrials.gov 中可以获得的临床研究资料有: 临床研究方案摘要, 它包括研究目的概述、参与者招募情况、患者参与标准、研究地点和具体联系方式; 其他信息包括研究设计、研究阶段、研究中使用的药物或疗法及用于何种疾病。并与包括 MEDLINEplus 和 PubMed 在内的医药网站中的相关疾病主题链接。

(三) NIST 的 Chemistry WebBook

是美国国家标准与技术研究院(National Institute of Standards and Technology NIST)的基于 Web 的物性数据库(<http://webbook.nist.gov/chemistry/>), Chemistry WebBook 可以看作是 NIST 的标准参考数据库 Standard Reference Data 中一部分与化学有关的数据库的 Web 版本,可通过分子式检索、化学名检索、CAS 登录号检索、离子能检索、电子亲和力检索、质子亲和力检索、酸度检索、表面活化能检索、振动能检索、电子能级别检索、结构检索、分子量检索和作者检索等方法,得到气相热化学数据、浓缩相热化学数据、相变数据、反应热化学数据、气相离子能数据、离子聚合数据、气相 IR 色谱、质谱、UV/Vis 色谱、振动及电子色谱等。

(四) Thomson Pharma 在研与上市药物情报数据库

Thomson Pharma(<http://www.thomson-pharma.com/>)是面向制药和生物技术公司的动态信息化综合医药信息平台。它的信息来源于各种科技文献、专利文件、新闻、科技会议论文以及金融方面的最前沿的信息等,相关信息包括药物、基因序列、化学、专利、商业、文献新闻等。

Thomson Pharma 以报告的方式提供信息。它将平台下数十个数据库提供的丰富信息整理为七大报告:①药物:26 000 多个药物(每月增加约 150 种药物);②公司:7500 家公司信息;③专利:超过 40 个专利授权机构的专利情报;④期刊和会议:来自 11 000 种同行评议的科技期刊文献和重要会议的信息;⑤化合物:超过三百万的化合物(每月增加约 5000 个化合物);⑥基因序列:超过八百万条(每月增加约 16 000~20 000 条);⑦靶标:5200 多个靶标的信息。其中每一份报告都由工作人员围绕主题将所有相关信息进行了整理和综合,并附有业界专家撰写的综述和评论。

其药物报告信息包括:药物概要,包括药物的名称,研发、上市状态,FDA 批准情况,治疗范围,作用方式等信息;药物相关的最新新闻;药物销售和市场情况的图示和描述,包括对将来几年的销售和市场占有情况的预测;药物在美国处方中使用情况的统计数据;药物在相应治疗领域内的应用情况;药物的研发历史和现状;药物注册申请情况;药物相关专利列表和超链接;业界专家根据 Thomson Pharma 收集的所有相关文献撰写的文献评论;从文献中摘录的科学数据,如结构-活性关系(SAR)的概要信息、ADMET 数据(吸收、分布、代谢、排泄和毒性)等化学、生物、代谢、临床等各方面的研究数据。

(五) Dialog 系统的药学信息

Dialog 系统是世界上最大的联机数据库检索系统(<http://www.dialog1.com/pharmaceuticals>),其药物信息分为新药要点(Drug Pipeline Summaries)、产业情报(Industry Intelligence)、工业通讯(Industry Newsletters)、医药专利信息(Pharma Patent Information)、监管信息(Regulatory Information)、研究文献(Research Literature)、研究新闻(Research News)等几大类,选中某类后进入检索状态。

(六) 药物专利检索

专利是新药开发和研制中最为重要的文献,药物专利包括药物用途专利、制备方法专利和药品专利(药物化合物专利、药物组合物专利、药物制剂专利)。有关专利检索请参见第七章第三节,从以下这些网站可获得有价值的药物专利信息:

(1) 各国药品管理机构网站:如美国 FDA 橙皮书。

(2) 各国专利局官方网站:如 SIPO、USPTO 等。

(3) 公益性免费专利数据库网站:如 free patents online(目前世界上唯一提供化学结构检索的免费专利数据库网站)。

(4) 商业性专利数据库网站:如 Delphion。

(5) 大型国际商业性联机检索系统:如 STN、DIALOG 等。

检索和使用药物专利时应注意以下问题:



(1) 检索制备方法专利: 除了检索制备方法外, 要对化合物专利进行全面检索, 规避侵权; 药物组合物专利应对药物组合物和制剂专利同时检索; 药物剂型专利, 除了特定剂型外, 应对关于该药物的剂型专利全面检索, 确保技术的竞争力。

(2) 保护区域的确定: 应进行全面的同族专利和法律状态检索, 确定专利保护的地理空白区。

(3) 使用号码途径利用优先权号检索同族专利: 要注意号码转换。如:

CN85108115 新低血糖剂的制备方法

优先权: 1984.11.13 US 670, 776

申请人: 美国伊莱利公司

利用欧洲 esp@cenet 网站检索, 则需将优先权号码转换成 US19840670776。

(4) 使用名称途径检索: 要注意同一公司的不同称谓。如美国保洁公司, 还普罗格特和甘保尔公司、普罗格特-甘保尔公司、普罗格特和甘布尔公司等名称。

(5) 使用主题途径: 要注意名词同义词、单复数及动词不同时代。如: 诺氟沙星(norfloxacin), 也叫氟哌酸, 淋克星, 力醇罗, 喹氟沙星, 正氟氯霉素, 1-乙基-6-氟-1, 4-二氢-4-氧代-7-(1-哌嗪基)-3-喹啉羧酸, 诺氟哌酸, 1-乙基-6-氟-4-氧代-1, 4-二氢-7-(1-哌嗪基)-3-喹啉羧酸等。

(6) 使用分类途径要分清国际专利分类法(IPC)、欧洲专利分类法(ECLA)、美国专利分类法(USPC)、日本专利分类法(FI/F-TERM)、中国专利范畴分类等的异同。

四、机构信息

(一) 药学机构与组织

医药院校和各国的药学会是药学机构与组织中最重要资源, 通过这些网站可以得到很多有用的信息, 如学校介绍、招生计划、工作招聘、学会机关刊物全文、继续教育信息、学术会议资料等。下面列出几个主要的站点。

1. American Society for Clinical Pharmacology and Therapeutics (美国临床药理学和治疗学学会, ASCPT): <http://www.ascpt.org>。

2. American Association of Colleges of Pharmacy (美国药学院学会, AACP): <http://www.aacp.org>。

3. American Society for Pharmacology and Experimental Therapeutics (美国药理学和实验治疗学学会, ASPET): <http://www.aspet.org>。

4. American Society of Health-System Pharmacists (美国医院药师学会, ASHP): <http://www.ashp.org>。

5. American Association of Pharmaceutical Scientists (美国药师学会, AAPS): <http://www.aapspharmaceutica.com>。

6. American Pharmacists Association (美国药剂师协会): <http://www.aphanet.org/>。

7. 中国执业药师协会: <http://www.clponline.cn/>。

8. 欧洲制药学联盟(EUFEPS): <http://www.eufeps.org/>。

9. 日本药学会: http://www.pharm.or.jp/index_e.html。

10. 国家食品药品监督管理局: <http://www.sda.gov.cn/>。

(二) 企业信息查询

1. Pharma-info 医药公司信息 (<http://www.pharma-info.com/en/index.html>) 该网站提供全世界 90 多个国家近 8000 个医药公司 (pharmaceutical companies) 和 100 个合同承包制造商 (contract manufacturers) 的商务信息, 包括公司地址、联系方式、业务范围等, 合同承包制造商栏提供了比较详细的企业可以加工的产品信息, 如片剂、胶囊、糖衣片或是产品包装信息等。同

时,还提供供需合作信息。制药企业(公司)还可以将自己的相关信息添加到其系统中。

2. 医药网(<http://www.pharmnet.com.cn>) 由浙江网盛生意宝股份有限公司创建于1999年,是目前国内最大最早的医药类综合网站之一。该网站的“医药企业”栏目提供了近3万家药品、医疗器械、保健食品、仪器设备、药品包装材料、辅料及其他相关企业的公司详细信息和产品信息。

五、综合性药学专业网站

(一) 中国医药信息网

中国医药信息网(<http://www.cpi.gov.cn/>)由国家药品监督管理局及其信息中心主办。主要为全国医药行业和药品监督机构提供各类医药市场、经济、管理、技术信息。该网站最有特色的是提供国内多种医药信息数据库检索服务(图8-43),包括以下几类:

1. 产品与企业类 基本药物库、药品包装材料库、国产药品库、保健食品库、经营企业库、生产企业新库、原料药企业库、进口药品库、生产企业库。
2. 市场与研发类 医药市场研究与分析库、价格库、国内药物研发信息库、医药生物制品库、制药原料及中间体库。
3. 国外信息类 世界批准新药库、国外药讯库、美国新药研究开发库、国外药讯2010—2014年(子库)。
4. 知识产权类 中药保护品种库、美国专利失效库。
5. 文献信息类 药学文摘库。
6. 政策信息类 中药法规库、政策法规库。
7. 其他 GMP企业、药品说明书库。



图 8-43 中国医药信息网数据库检索主页

(二) Medscape

Medscape (<http://www.medscape.com>) 是一个综合性的医学网站, 按照专业提供信息。可按以下 3 种途径查询药物信息:

- (1) 选择其主页的 Drug Reference 检索选项, 输入药名即可查找药物信息。
- (2) 利用 <http://www.medscape.com/druginfo/onotype=monograph> 检索页面进行高级检索(可选择药名或适应证)。
- (3) 从药师(Pharmacists, <http://www.medscape.com/pharmacists>) 专业入口查找与药物相关的信息。

Medscape 的药物信息数据库提供了详细的药品信息, 包括药品的分类、用法和剂量、理化属性、药理毒理、患者使用手册、副作用、用药注意事项以及生产厂商和商品名等。

(三) FDA 的药物信息

FDA (<http://www.fda.gov>) 是美国食品药品监督管理局的官方网站, 具有一定的权威性。主要信息包括应急准备, 药品认证和数据库, 药品安全和可用性, 药品开发和审批程序, 指导、一致性和监管信息, 新闻动态, 药物科学与研究等。这些资源针对不同的用户群体(消费者、卫生保健专业人士、药物研制者)分类提供(Resources for You), 其中, 通过等效性评估获得批准的药物(橙皮书)可为药物仿制提供依据。

针对卫生保健专业人士的药物信息(Information for Healthcare Professionals)对临床医生有很大帮助(图 8-44)该栏目可以查找以下信息:

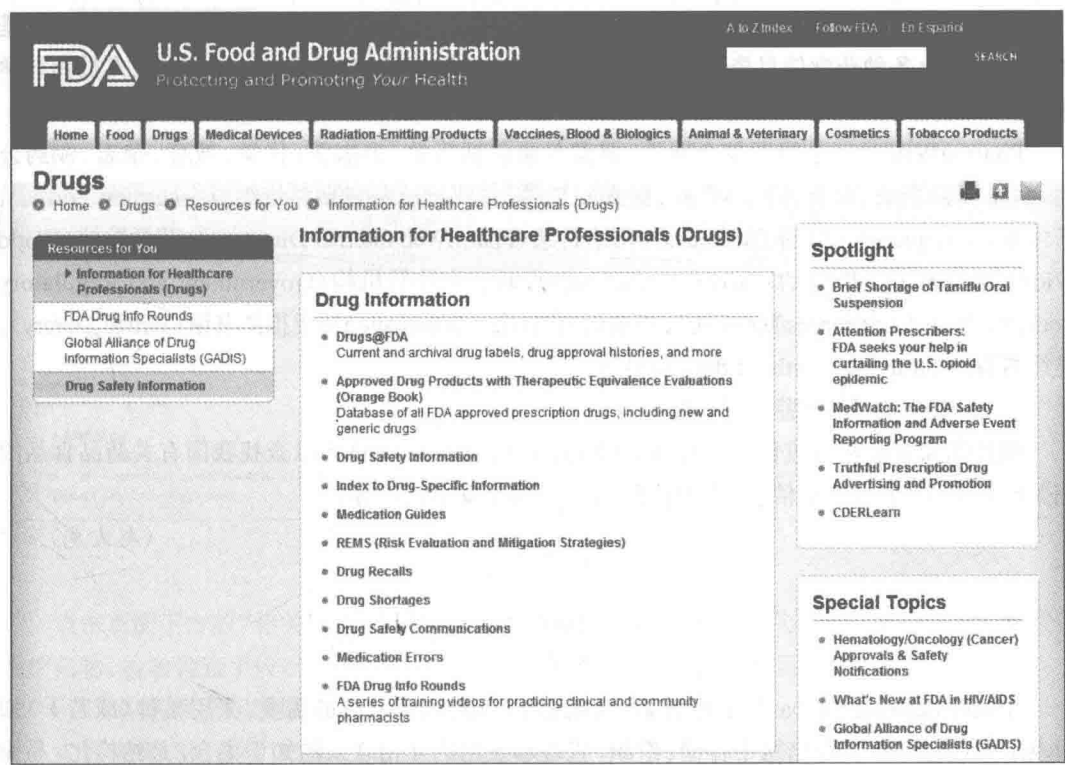


图 8-44 FDA 针对卫生保健专业人士提供的药物信息主页

- 1. Drugs@FDA 可以检索当前药品和历史药品的标签和申请的历史档案。
- 2. 通过等效性评估获得批准的药物(橙皮书) 收录的是美国 FDA 根据安全与有效原则收录的所有批准药物, 包括新药、仿制药。橙皮书将所有药品按照治疗等同性分类以方便医务人员参考; 同时提供专利和保护信息, 为仿制药申请提供依据。可通过活性成分检索(Search by

Active Ingredient)、申请人检索(Search by Applicant Holder)、商品名检索(Search by Proprietary Name)、申请号检索(Search by Application Number)、专利号检索(Search by Patent)五个入口进行检索:选择某一入口点击进入后,还可选择处方药(Rx, Prescription Drug Products)、非处方药(OTC, Over-the-Counter Drug Products)、停用药(Disc, Discontinued Drug Products)进行检索。

3. 药物安全信息。

4. 药物特殊信息索引 专业人士早期对部分有关药物的评价。

5. 药物短缺信息。

6. 用药错误信息。

(四) 虚拟药学图书馆

虚拟药学图书馆是综合性的药学资源指南目录(<http://www.pharmacy.org/>),包括药学院校、制药公司、期刊和图书、数据库、政府机构、学会协会、学术会议、新闻组、医院等。在 Pharmacy Schools 栏提供了全世界数十个国家的药学院校链接, Journals and Books 栏提供了上百种杂志的链接。该网站通过其提供的网络链接提交(Add/Correct a Link?)功能,允许访问者提交或修改相关站点的 URL,增加了用户的参与性。

(五) 药物网

药物网(<http://www.drugs.com/>)是得到 HON 认证的药学网站,其信息可靠。其药物信息主要来源于 Wolters Kluwer Health, Physicians' Desk Reference, Cerner Multum 和 Thomson Micromedex 的相关信息。

(六) Pharmaweb

PharmaWeb(<http://www.pharmweb.net>)是第一个在因特网上提供药学信息的专业网站,基本涵盖了网上各种药学信息资源。该网站拥有庞大的用户群,包括患者、卫生专业人员和科学家,遍及世界一百五十多个国家。

PharmaWeb 是一个资源整合网站,资源目录包括药学、生物学、化学、教育、杂志、制药公司、医师药师名录、世界各国药学会、医药院校等,提供的服务包括讨论组(Discussion Forms)、新闻组(Newsgroups)、团体(Societies)、网络目录(PharmWeb Internet Directory)、药学院校(World Wide Pharmacy Colleges/Departments/Schools)、政府和监管机构(Government and Regulatory Bodies)、会议(Conferences/Meetings)、患者信息(Patient Information)、虚拟图书馆(Virtual Library)、继续教育(Continuing/Further Education)等。

(七) 国家食品药品监督管理局

通过国家食品药品监督管理局网站(<http://www.sda.gov.cn>)可以查找我国有关药品食品的相关政策文件、各种注册信息、审批信息、公告、药品安全信息等。

(赵文龙)

第六节 疾病知识库检索

知识库(knowledge base)是针对某一(或某些)领域问题求解的需要,采用某种(或若干)知识表示方式在计算机存储器中存储、组织、管理和使用的互相联系的知识集合,是结构化、易操作、易利用、全面有组织的知识集群。

医药类知识库国内以聚合 CMDK 为代表,国外以 First Consult 为代表,均是商业数据库,需要获得授权后才能使用。其数据均高度结构化,疾病有关的常见项目如病因、病机、诊断、治疗、预后等内容均以字段形式结构化展示。

临床医药知识库或医学百科常常作为患者教育资源在医药卫生普及网站免费向公众提供服务。免费的医药知识库国内以百度百科为代表,国外推荐使用 Medscape 的 Reference 系统和



MedlinePlus 的 Health topics, 医学维基百科 (medpedia) 也有很高的权威性, 但其检索功能不是很强。

一、中文疾病知识库

(一) 临床医药学知识互动平台

临床医药学知识互动平台 (<http://cmkd.juhe.com.cn/>) 是目前国内比较系统提供临床疾病及相关基础信息的知识库系统, 通过该系统可以获取常见疾病的流行病学、病因、发病机制、临床表现、并发症、实验室检查、辅助检查、诊断、鉴别诊断、治疗、预后以及预防等内容。

检索步骤:

1. 在快速检索状态输入“唐氏综合征”, 点击【快速检索】按钮 (图 8-45);
2. 获得结果 1 条, 点击标题名称“小儿唐氏综合征” (图 8-46);
3. 详细显示检索结果相关内容 (图 8-47)。通过点击显示页面左侧结构化的条目查看相关信息, 如流行病学查看该病的发病率, 发病机制查看致病的原因和发病机制等。

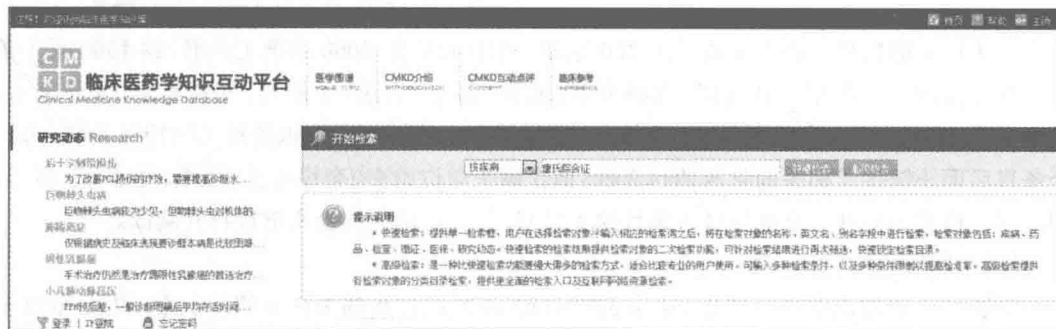


图 8-45 CMKD 快速检索界面

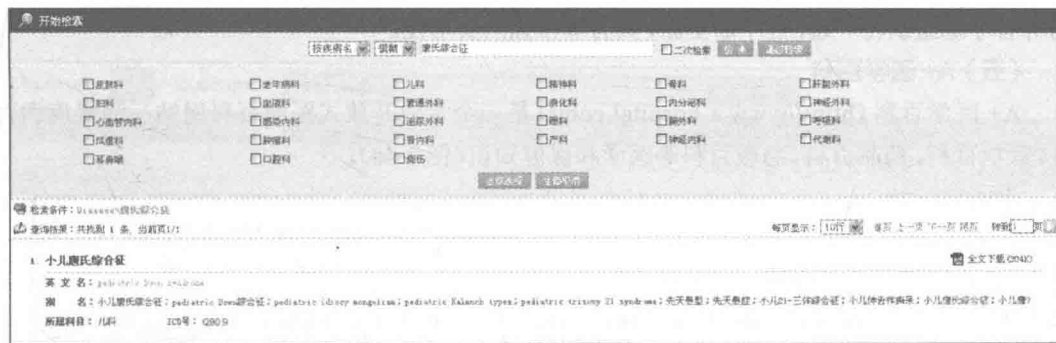


图 8-46 CMKD 结果显示界面

点击页面下方的“相关药品”、“相关检查”、“相关循证”、“研究动态”按钮查看与该病相关的治疗药物、检验检查手段、循证研究成果和期刊杂志有关该疾病的最新研究文献。

(二) 中国疾病知识总库

中国疾病知识总库 (China Disease Knowledge Total Database), 简称疾病库 (China Disease Database, CDD) (http://www.qzygz.com: 8082/CDD/Disease/Dis_Basic.aspx), 由解放军医学图书馆研发的一个面向临床医药学专业人士, 同时兼顾大众的知识型数据库。包括疾病数据库、药品数据库、辅助检查数据库和循证医学数据库, 各数据库相互关联, 实现知识链接与漫游, 形成网状的知识体系。字段结构和使用方法与 CMKD 相似。

(三) CNKI 知识库

通过 CNKI 的知识搜索的工具书库 (http://gongjushu.cnki.net/refbook/de_fault.aspx) 也可以

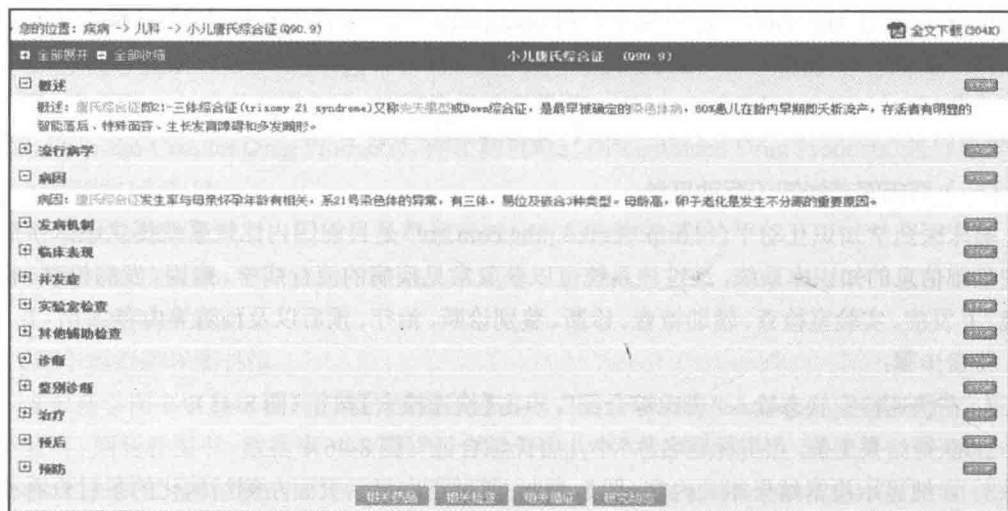


图 8-47 CMKD 条目详细结果显示界面

检索有关疾病的知识。该库集成了近 200 家知名出版社的近 4000 余部工具书,约 1500 万个条目,70 万张图片。类型包括词典、百科全书、表谱、传记、语录、手册等,涵盖哲学、文学艺术、社会科学、文化教育、自然科学、工程技术、医学等各个领域。各知识条目之间相互关联,每一个条目后面还链接了相关的学术期刊文献、博士硕士学位论文、会议论文、报纸、年鉴、专利、知识元等。检索结果列出全部含有该条目的工具书,通过其提供的全文链接打开阅读。

(四) 医脉通诊疗知识库

医脉通诊疗知识库 (<http://disease.medlive.cn/wiki/>) 是以疾病为核心的知识库,收录 7000 多种疾病,遵循医生诊疗的实际过程,每一种疾病按照统一的标准框架组织内容,涵盖疾病的预防、诊断、治疗、预后整个过程,把和某种具体的疾病相关的最新研究证据、指南、病例、文献资料等有序地组织在一起,且不断更新,具有非常强的实用性。

(五) A+ 医学百科

A+ 医学百科 (<http://www.a-hospital.com/>) 是一个专业开放式医学百科网站。涵盖疾病百科、症状百科、药品百科、急救百科等医学和保健知识(图 8-48)。



图 8-48 A+ 医学百科疾病百科知识条目

（六）百度百科

目前有医学类条目近 3.6 万多个(<http://baike.baidu.com>)。任何人均可参与其词条的任何段落的编辑和完善,与维基百科类似。词条与词条之间经纬交错、相互关联。

二、国外疾病知识库

（一）First Consult

首席参考(First Consult)是 Elsevier Science 出版发行的针对临床医生提供的临床知识库,以循证医学知识为核心内容,以精选的高质量期刊、临床指南和 FDA 药物信息作为补充,每个条目均采用高度结构化的文本形式展示。在现阶段,First Consult 提供了及时的、人性化的关于评估、诊断、临床管理、预后和预防的最新信息。

可通过 Clinicalkey (<http://www.clinicalkey.com>) 和 MD Consult (<http://www.mdconsult.com>) 网站访问。First Consult 同时提供 PDA 版本。具有以下特点:

1. 数据库的信息基于临床主题,并且不断更新;
2. 其操作程序以生动的视频和动画展示;
3. 临床资料都基于循证;
4. 提供鉴别诊断工具;
5. 有 1 万个以上可定制的患者教育讲义;
6. 综合药物数据库提供准确的 FDA(美国食品药品监督管理局)药品说明书;
7. 界面简洁、疾病分类清晰,检索方便;
8. 除与临床相关的信息,还提供医学植物介绍、小测试等;

本节以 ClinicalKey 为入口网站检索。进入 ClinicalKey 网站,在检索输入框中输入检索词,如“Down syndrome”,点击【search】按钮,返回结果如图 8-49,选择左侧导航栏“First Consult”。

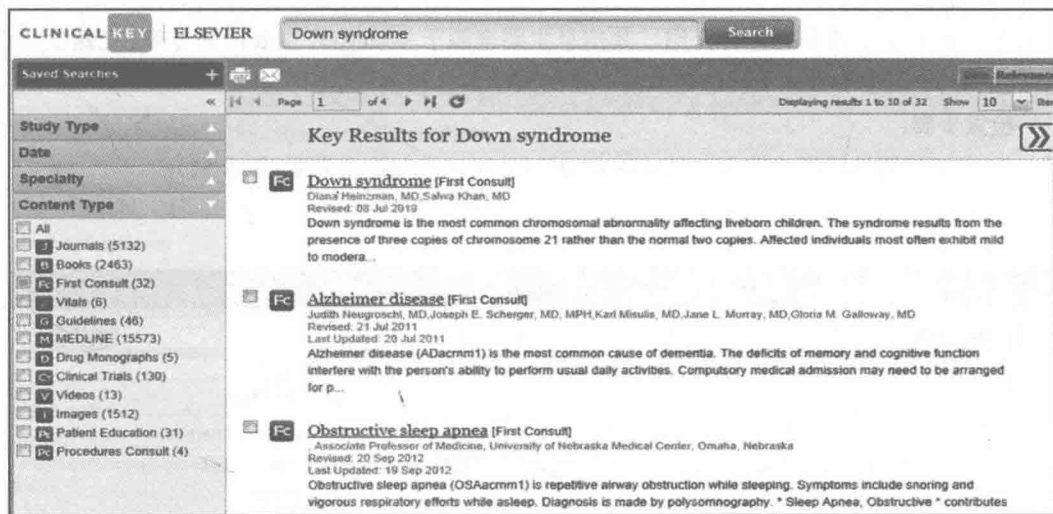


图 8-49 ClinicalKey 检索结果

First Consult 条目的详细信息采用结构化文本显示,包括概要、背景(病因、流行病学、ICD 编码)、诊断(症状、体征、相关疾病、鉴别诊断)、治疗(治疗选项、随访、合并用药/疾病、患者的满意度和生活方式的优先权)、预后、预防、资源(参考文献、学会)等(图 8-50)。

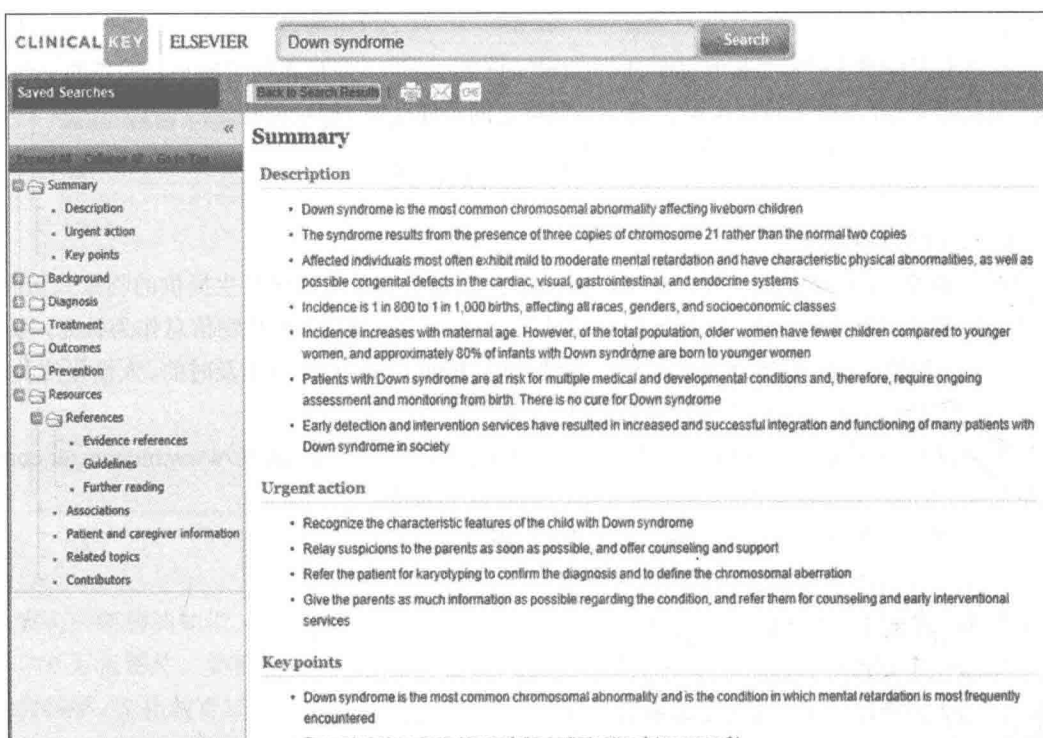


图 8-50 First Consult 条目结构化详细信息

(二) 医学网景疾病参考资料

医学网景(Medscape)网站参考资料(Reference)下的疾病和状态(Diseases& Conditions)栏目所提供的与疾病相关的信息分为内科(Medicine)、外科(Surgery)和儿科(Pediatrics)三个主题领域,可通过浏览查询,其词条不仅仅是疾病,也有和疾病相关的如药物中毒、临床操作等。每个词条都署名了作者并有作者介绍。其信息主要来源于 WebMD, 主要面对专家、初级卫生保健医生和其它卫生专业人士提供集成的丰富的医学信息和教育学习工具。

检索步骤:

1. 在浏览器地址栏内输入 reference.medscape.com 进入检索界面(图 8-51);

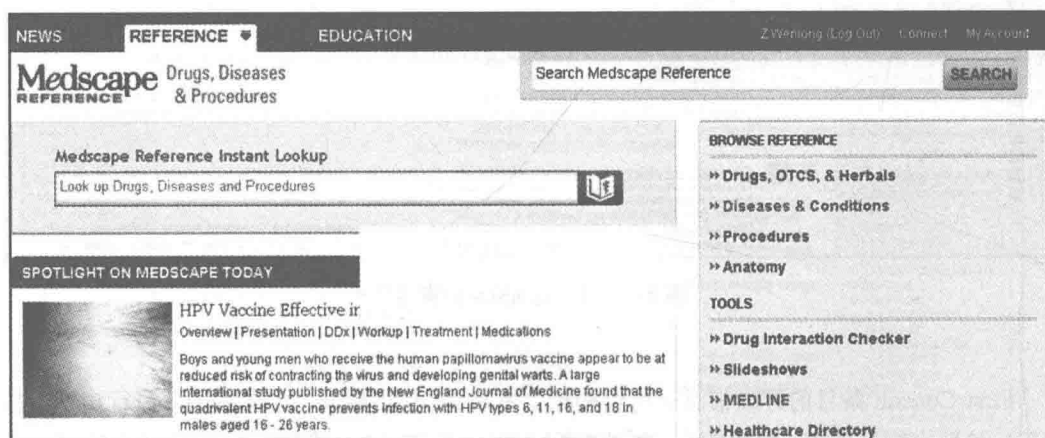


图 8-51 Medscape 医学参考库主页

2. 在检索输入框中输入 Down syndrome, 系统将在医学参考库各栏目中的信息直接显示出来(图 8-52);

Search Results in Drugs & Diseases (1199)

Filter Results

Conditions (959)

- Adult (766)
- Pediatric (193)

Procedures (201)

Drugs (34)

Clinical Case (5)

Pathology of Myeloid Proliferations Related to Down Syndrome

Pathology of Myeloid Proliferations Related to Down Syndrome. For individuals with Down syndrome (DS), the risk of developing acute leukemia is increased ...

Ophthalmologic Manifestations of Down Syndrome

Ophthalmologic Manifestations of Down Syndrome. About 60% of patients with Down syndrome have ophthalmic manifestations. ... with Down syndrome ...

Prenatal Imaging Findings in Down Syndrome

Prenatal Imaging Findings in Down Syndrome. Down syndrome is a relatively common congenital disorder caused by the presence of an extra 21st chromosome.

Alzheimer Disease in Down Syndrome

Alzheimer Disease in Down Syndrome. Alzheimer disease (AD) is the most common form of dementia. ... Individuals with Down syndrome (DS), or ...

Down Syndrome

Overview | Presentation | DDx | Workup | Treatment | Medication

Down Syndrome. Down syndrome is by far the most common and best known chromosomal disorder in humans and the most common cause of intellectual disability.

LATEST NEWS RESULTS FOR DOWN SYNDROME

- Thai Authorities to Probe Surrogacy Clinics Amid Down Case Controversy August 2014
- Dosing the New Oral Anticoagulants in Atrial Fib: Can We Do Better? August 2014
- Lower LDL Still Best: Very Low Levels of LDL Linked With Reduced CV Events July 2014

ADVERTISEMENT

Missing your sales rep?

Read about the newest product information from Industry

READ MORE



图 8-52 Medscape 检索结果列表

3. 选择 Down Syndrome, 打开查看详细内容。Medscape 有关疾病的知识, 也采用高度结构化的文本通过标签形式展示(图 8-53)。包括概述(Overview)、表征(Presentation)、鉴别诊断(DDx)、检查(Workup)、治疗(Treatment)、药物(Medication)和随访(Follow up)等。

Down Syndrome

Author: Harold Chen, MD, MS, FAAP, FACMG; Chief Editor: Bruce Buehler, MD more...

Print

Share

Feedback

Overview Presentation DDx Workup Treatment Medication

Updated: Apr 10, 2014

Practice Essentials

Background

Pathophysiology

Etiology

Epidemiology

Prognosis

Patient Education

Show All

Multimedia Library

References

Practice Essentials

Down syndrome is by far the most common and best known chromosomal disorder in humans and the most common cause of intellectual disability. It is primarily caused by trisomy of chromosome 21, which gives rise to multiple systemic complications as part of the syndrome. However, not all defects occur in each patient; there is a wide range of phenotypic variation.

Signs and symptoms

When recording the history from the parents of a child with Down syndrome, the clinician should include the following:

- Parental concern about hearing, vision, developmental delay, respiratory infections, and other problems
- Feeding history to ensure adequate caloric intake
- Prenatal diagnosis of Down syndrome
- Vomiting secondary to gastrointestinal tract blockage by duodenal web or atresia
- Absence of stools secondary to Hirschsprung disease
- Delay in cognitive abilities, motor development, language development (specifically expressive skills), and social competence
- Arrhythmia, fainting episodes, palpitations, or chest pain secondary to a heart lesion
- Symptoms of sleep apnea, including snoring, restlessness during sleep, difficulty awaking, daytime somnolence, behavioral changes, and school problems

图 8-53 Medscape 检索结果详细显示

Notes

(三) Medline Plus

由 NLM 提供 (<http://www.nlm.nih.gov/medlineplus/>), 其健康专题 (Health Topics) 栏目提供了 800 多个疾病有关的症状、病因、治疗和预防以及个人健康状况和心理健康评估方面的内容 (图 8-54), 并定期更新。



图 8-54 MedlinePlus 健康主题引导页面

A.D.A.M 医学百科全书 (Medical Encyclopedia, http://www.nlm.nih.gov/medline_plus/encyclopedia.html), 可查询关于疾病、试验、检查、症状、损伤、外科等相关条目 4000 多个。

(四) 医学维基百科

由经过个人申请并审核批准的、来自著名医学院或者研究机构 (如哈佛医学院、斯坦福大学医学院等) 的专家学者编写 (<http://wiki.medpedia.com/>)。它是一个开放式的医学百科, 和维基百科不同的是, 访问者不能对词条文章进行修改。每一个词条都附有作者的资料。

点击浏览全部文章 (See All Articles), 进入按字母顺序查询页面, 如查询“唐氏综合征”, 可单击字母 D, 浏览找到 Down syndrome, 打开链接阅读。

(五) 网络医学词典检索

包括大量的医学词典 (<http://www.kmle.com/>), 如 MeSH、ICD-10、Dorlands 医学词典、Merriam-Webster 词典、Drugs.com 的药物、WebMD、医学缩略语等。

(赵文龙)

第七节 医学视频数据库检索

一、万方医学视频数据库

万方医学视频数据库系统 (<http://v.med.wanfangdata.com.cn>) 是以医学视频为主要内容的知识服务系统, 并与中国科技信息研究所、中华医学会、中医药管理局进行广泛的合作, 收录的栏目包括《中华医学会继续教育视听杂志》、《中华医学会继续教育视听教材》、《全科医生系列》、《名医谈百病系列》等多个系列的 900 余部医学视频资源 (图 8-55)。该视频资源库可与万方医学网中相关期刊论文、学术论文、会议论文文献进行相互链接, 方便用户直接追踪文献。





图 8-55 万方医学视频数据库

二、Primal Pictures 人体解剖 3D 模型数据库

Primal Pictures 三维互动式立体解剖学数据库利用真实医学数据——CT 扫描、MRI 等进行精确的三维模型重建，全面汇集了超过 6500 个偏重于独立器官、身体部位或解剖系统的高精度三维动态互动式解剖模型。不论是临床应用、康复治疗、教学或研究中的运用，也不论是专业或非专业的人士，均可以通过数据库直接进入三维动画，查看精确的人体解剖模型，详细了解其结构与功能、生物学特性及治疗与手术操作过程等信息。

Primal Pictures 三维互动式立体解剖学数据库可实现以下功能：①剥离解剖模型组织层，可旋转 360°，并将其已剥离的组织层恢复至其初始结构形态；②将 3D 模型与 MRI 图像进行对比，利用交互观测对一些伤病进行诊断与管理；③通过专业的文字描述及解剖标本和病理幻灯片来揭示临床解剖的相关性；④查看临床视频，更好地了解手术操作流程和治疗方式；⑤通过显示肌肉、韧带和骨骼之间相互作用的 3D 动画演示运动功能；⑥方便地将图像、动画、视频短片、幻灯片及其补充的文字描述进行关联；⑦可将文本、图像、动画、视频短片和幻灯片从窗口直接保存以备使用；⑧内嵌学习指导与针对性练习题，方便进行自我评估测试，强化学习效果。

Primal Pictures 三维互动式立体解剖学数据库主要包含了以下 6 个模块：局部解剖 (Regional Anatomy)、系统解剖 (Systemic Anatomy)、运动创伤与治疗 (Sports & therapy)、特殊专题 (Speciality titles)、外科手术 (Surgery) 和三维实时人体模型 (3D Real-time Viewers)。

三、美国视频实验期刊 (JoVE)

美国 Journal of Visualized Experiments (JoVE) 视频实验期刊创刊于 2006 年，是全球第一个基于视频的同行评议实验期刊，并被 PubMed 所收录。

JoVE 的期刊专注于发表生物学、医学、化学和物理学领域的视频资源。与传统文本格式期

刊相比, JoVE 突出的特色在于综合了多媒体的优势, 运用视频使知识传递更加生动直观, 也在很大程度上解决了复杂生物、医学实验难以还原和重复进行的问题。借助这一新型互联网出版模式, JoVE 帮助全球的科研工作者节省宝贵的时间和精力, 足不出户可抵达全球顶级实验室。该数据库每日更新, 每月新增 50 项实验视频资源。

四、爱思唯尔 ClinicalKey 医学视频库

ClinicalKey (简称 CK) 是 Elsevier 基于全部医学内容开发, 2012 年全新推出的全医学平台。ClinicalKey 资源库包括电子期刊全文、电子书、年鉴、视频、图片、药物专论、临床试验、诊疗指南、循证医学等 12 类的文献资源, 同时建立了自身的医学分类系统, 旨在帮助用户用最短的时间、最简洁的方式找到最想要的资源。

ClinicalKey 资源库之一的 Procedures Consult 是一个强大的在线操作视频参考工具, 它提供简明的访问方式来学习当今医院最常用的必备医疗操作的准备、执行和落实等完整细节 (图 8-56)。Procedures Consult 包括了 ABIM (美国内科学委员会) 要求必修的医疗操作并遵循 ACGME 的标准, 以帮助医生和住院医师取得最好的治疗效果。



图 8-56 Procedures Consult 在线操作视频库

Procedures Consult 提供七个专科的操作视频及系统的操作相关信息, 包括麻醉操作 (34 种), 急救医学操作 (73 种), 家庭医学操作 (51 种), 普通外科操作 (47 种), 内科操作 (46 种), 骨科操作 (52 种), 医生培训操作 (34 种)。

本章小结

本章主要介绍中外主要学位论文、会议文献、专利文献及药学相关数据库的收录内容与检索方法, 并重点以 Cochrane library 为例介绍循证文献的类型以及常用循证医学文献检索系统的使用, 此外, 本章也简要介绍了万方医学视频、Primal Pictures 3D 解剖数据库等常用医学教学与临床视频资源。



思考题

1. 通过哪些数据库可以查找到国内学位论文?
2. 如何通过网上获取某一学术专题(如高血压)的国内外学术会议预告?
3. 举例说明, 如何通过会议注册系统, 提交参会论文。
4. 利用“医生指南 - 会议资源中心”查找在北京举办的医学会议信息。有哪些途径?
5. 利用“HON 会议信息”查找有关神经系统疾病(Nervous System Diseases)的会议信息。
分别用检索和按学科分类(Browse)浏览两种途径查找, 并比较两种检索结果的不同。
6. 循证医学的证据按加工的深度不同, 可分为哪 5 类?
7. 循证医学的证据的核心资源有哪些?
8. 利用 USPTO、Espacenet(四个专利库)、WIPO、free patents online 专利检索系统检索有关聚焦超声的授权专利, 比较说明各专利检索系统的特点。
9. 使用 Rxlist、Toxnet、PDR、PubChem、MedWatch 检索有关达菲(奥司他韦, Oseltamivir)的处方资料、研究文献(新药报批、临床研究)、不良反应原始报道等。
10. 利用 NGC、EBMR 检索妊娠期高血压的治疗指南。

第九章 生物信息数据库检索

生物信息学是一门新兴的交叉学科,涉及生物信息的获取、管理、存贮、分析、解释和应用等各个方面,综合运用数学、计算机科学和生物学等学科的知识与方法,来阐明和理解大量生物数据所包含的生物学意义。生物信息数据库是生物信息学的重要组成部分,是非常重要的生物信息学资源。随着生物信息学的快速发展,生物信息数据库的种类和数据量日益增多,数据库功能日益完善,在人类基因组计划、基因组学、生物信息学发展中起着非常重要的作用。

第一节 生物信息数据库概述

一、生物信息数据库的类型

生物信息数据库,是指在计算机存贮设备上合理存放且相互关联的生物信息集合。根据收录的信息内容主要分为以下五类:

1. 序列数据库 主要收录序列数据,它是最基本的生物信息数据库。根据序列分子类型不同,又可分为:DNA 序列数据库,如 GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>)、ENA (<http://www.ebi.ac.uk/ena/>)、DDBJ (<http://www.ddbj.nig.ac.jp/>)等;RNA 序列数据库,如非编码 RNA 数据库(the Non-coding RNA Database, ncRNAdb, <http://biobases.ibch.poznan.pl/ncRNA/>),微小 RNA 数据库(the MicroRNA Database, miRBase, <http://www.mirbase.org/>)等;蛋白序列数据库,如通用蛋白质资源(the Universal Protein Resource, UniProt, <http://www.uniprot.org>)等;基因组序列数据库,如酵母基因组数据库(the Saccharomyces Genome Database, SGD, <http://www.yeastgenome.org/>)、鼠基因组信息(Mouse Genome Informatics, MGI, <http://www.informatics.jax.org/>)等。

2. 结构数据库 主要收录蛋白质、多肽的三维结构(X线和磁共振测定的)数据以及酶、病毒、碳水化合物和核酸的晶体结构数据库,如蛋白质结构数据库(the Protein Protein Data Bank, PDB, <http://www.rcsb.org/pdb/home/home.do>)、核酸结构数据库(the Nucleic nNucleic Acid Database, NDB, <http://ndbserver.rutgers.edu/>)、剑桥结构数据库(Cambridge Structural Database, CSD, <http://www.ccdc.cam.ac.uk/Solutions/CSDSystem/Pages/CSD.aspx>)等。

3. 图谱数据库 主要收录基因组图谱数据,如遗传图谱、物理图谱、转录图谱和序列图谱,如 NCBI 的基因组图谱(<http://www.ncbi.nlm.nih.gov/genome>)、EBI 的 Ensembl (<http://www.ensembl.org/index.html>)、UCSC 基因组浏览器数据库(UCSC Genome Browser Database, <http://genome.ucsc.edu/>)等。

4. 突变数据库 主要收录基因突变以及多态性数据,有综合性和特殊位点突变数据库两种。综合性突变数据库收录多种类型的突变或多态性数据,如单核苷酸多态性数据库(Single Nucleotide Polymorphism Database, dbSNP, <http://www.ncbi.nlm.nih.gov/projects/SNP/>)、人类基因突变数据库(Human Gene Mutation Database, HGMD, <http://www.hgmd.org/>)等;特殊位点数据库主要收录某一位点或某一特殊类型的基因突变和多态性数据,如人类肿瘤 P53 突变数据

库(p53 Mutation in Human Cancer, <http://p53.free.fr/>)、APC 基因突变数据库(<http://www.umd.be/APC/>)等。

5. 文献数据库 主要收录各种与生物信息有关的文献,如 PubMed(<http://www.ncbi.nlm.nih.gov/pubmed>)、在线人类孟德尔遗传数据库(Online Mendelian Inheritance in Man, OMIM, <http://www.omim.org/>)等。

二、生物信息数据收集与存储

(一) 生物信息数据的收集

1. 人工收集 一些主要的生物信息数据库如 GenBank、ENA、PIR、SWISS-PROT, 在建库的初期主要靠人工搜索科学期刊中核酸和蛋白质序列数据,然后录入到数据库中。这种收集方式费时费力,跟不上生物信息数据增长的速度。

2. 直接发送 随着测序技术和克隆技术的发明和广泛应用,使得序列数据急剧增长,特别是人类基因组和各种模式生物基因组计划的启动和顺利实施,生物信息数据更是呈指数增长,巨大的生物信息数据仅靠科学期刊出版发行已经不可能,靠人工收集科学期刊中的信息数据更是不可能。因此,在 1988 年序列数据库与经常刊登序列数据的科学期刊合作,要求科研人员在论文发表之前必须将序列数据发送到某个数据库中,并从后者获得一个序列存取号(accession number),该存取号可随论文发表,代表该序列数据。因此,科研人员直接发送成为生物信息数据库收集数据的主要方式。

一些小型的二次数据库仍然采用电子邮件方式收集生物信息数据,但是大型核酸序列数据库如 GenBank、ENA、DDBJ,主要采用 Web 发送软件收集科研人员的数据,如 GenBank 的 BankIt(<https://www.ncbi.nlm.nih.gov/WebSub/?tool=genbank>)、ENA 的 Webin(<https://www.ebi.ac.uk/ena/submit/sequence-submission>)、DDBJ 的 NSSS(Nucleotide Sequence Submission System, <http://www.ddbj.nig.ac.jp/sub/websub-e.html>)等,一般需要注册才能使用。Web 发送软件因其快捷、方便、直观而成为科研人员发送数据的首选。

3. 成批发送 从 1998 年开始,人类基因组计划重心已从基因作图转向大规模基因组测序,一些专门的基因组测序中心,如 Sanger 和专门的商业性测序中心相继成立,大型制药公司为了争夺新药开发的制高点,不惜斥巨资投入人类基因组和病原生物基因组测序。这些测序中心和制药公司每天能测序几 Mb 或几百 Mb 碱基,一次需要发送大量的基因组数据,上述发送工具显然不能满足大规模发送序列数据的需要。为此,GenBank、ENA 和 DDBJ 均开发了大批量发送序列数据工具,如 NCBI 的 tbl2asn,ENA 的基因组组装发送(Genome Assembly Submissions)和 DDBJ 的集中发送系统(Mass Submission System, MSS)。因此,成批发送成为数据库收集数据的另一个主要方式。

4. 数据交换 数据库之间的数据交换也是生物信息数据库收集数据的一个重要方式,如 GenBank、ENA、DDBJ 三大核酸序列数据库建立国际核酸序列数据库协作体(International Nucleotide Sequence Database Collaboration, INSDC),分别收集所在区域的序列数据,实行数据共享,每天交换各自收集的序列数据。

(二) 生物信息数据的存贮

1. 记录格式 不同的生物信息数据库所采用的记录格式也不尽相同。由于 ENA 和 GenBank 是最为主要的核酸序列数据库,所以 ENA 格式和 GenBank 格式被其他生物信息数据库广泛采用。详见表 9-1。

(1) ENA 格式: EMBL 的每一条记录为纯文本文件,包括五个部分: general information(一般信息)、description(描述)、references(参考文献)、features(特征表)、sequence(序列),每一部分又分成几个子类(表 9-1)。其中“特征表”包含一批关键字(包括内含子、外显子、起始密

码子、基因及其功能等),它与 GenBank、DDBJ 是统一的,具体描述参见 <http://www.insdc.org/documents/feature-table>。

(2) GenBank 格式: GenBank 的每一条记录也为纯文本文件。每行左端为空格或标识符。标识符均为一完整英文字母,不用缩写。每一条 GenBank 记录,从 LOCUS 到 ORIGIN 为记录的注释部分,注释部分按标识符分成若干段。

表 9-1 ENA 格式和 GenBank 格式标识符的含义

ENA 标识符	GenBank 标识符	含义
General Information	LOCUS	记录名称、性质描述
Primary Accession	不显示	初始存取号
Accession	ACCESSION	序列存取号
SRS Entry ID	无	SRS 记录号
Molecule Type	在 LOCUS 部分	分子类型
Sequence Length	在 LOCUS 部分	序列长度
Entry Division	不显示	记录分区
Entry Data Class	在 LOCUS 部分	记录数据分类
Sequence Version	VERSION	序列版本号(AC 不变)
Creation Date	不显示	创建日期
Modification Date	在 LOCUS 部分	修改日期
Description	DEFINITION	序列简要描述
Keywords	KEYWORDS	关键词
Organism	SOURCE	来源种属
Organism Classification	ORGANISM	生物体分类
Comment	COMMENT	序列评述
References	REFERENCE	参考文献
Features	FEATURES	特征表
Key	source	关键字
Location	Location	定位
Qualifier	Qualifiers	限定
Value	直接列出	取值
Sequence (FASTA 格式)	ORIGIN	序列

2. 序列格式 不同的序列分析软件所采用的序列格式也不尽相同,因而存在多种不同的序列格式,其中最常见的序列格式是 FASTA 格式,又称 Pearson 格式(图 9-1),主要用于序列类似性检索或序列同源分析。

```
>gi|2582741|emb|AJ002508.1|HIVAJ2508 Human immunodeficiency virus type 1 protease gene
CCTCARRTCACTCTTTGGCARGACCCCTCGTCACAATAAAGATAGGGGGCAACTAAAGGAAGCTCTATTAGATACAGRAGCAGATG
ATACAGTAKTAGAAGAMATRASSTTTRCCAGGAAGRTGGAAACCAAAAATGATAGGGGAATTGGAGGTTTWTCAAGTAAGACAGTA
TGATCAGATACTCRTAGAAATCTGTTGACTCAGMTTG GTTGCACTYTAATTTT
```

图 9-1 FASTA 格式

第一行是说明行,以“>”符号开始,紧接着是序列存取号和描述性文字,有助于序列的理解。从第二行开始为序列本身。核苷酸符号大、小写均可,但氨基酸一般用大写字母。每一行都不得超过 80 个字母。由于 FASTA 格式无特殊的序列结束标志,建议最后多留一个空行。

三、生物信息数据库的查找

(一) 搜索引擎

搜索引擎是查找生物信息数据库的一个重要途径。利用生物信息数据库名称或部分信息,利用 Google、百度等搜索引擎进行查找(详见第十章第二节),获取该数据库的 URL 网址及相关信息,再到该数据库网页,获取其中的信息。

(二) 专门的生物信息数据库目录

从 2000 年开始, *Nucleic Acids Research* 设立了一个数据库目录 (http://www.oxfordjournals.org/our_journals/nar/database/c/)。该目录将所收录的 1000 多个数据库分成 Nucleotide Sequence Databases (核酸序列数据库)、RNA Sequence Databases (RNA 序列数据库)、Protein Sequence Databases (蛋白质序列数据库)、Structure Databases (结构数据库)、Genomics Databases (non-vertebrate) (基因组数据库(非脊椎动物))、Metabolic and Signaling Pathways (代谢与信号通路)、Human and Other Vertebrate Genomes (人类与其他脊椎动物基因组)、Human Genes and Diseases (人类基因与疾病)、Microarray Data and Other Gene Expression Databases (微阵列数据与其他基因表达数据库)、Proteomics Resources (蛋白质组资源) 等 14 大类,可按字顺或分类浏览,可以下载数据库介绍全文,并可以链接到所需要的数据库。

(三) 生物信息学中心资源导航

一些著名的生物信息学中心不仅自己建立和维护大量的生物信息数据库,而且一般在网上提供资源导航。例如,美国国家生物技术信息中心(NCBI)创建和维护的许多著名的生物信息数据库,如 GenBank、dbEST、dbGSS、dbSTS、dbSNP 等。欧洲生物信息研究所(EBI)的检索服务(EBI Search)提供了 90 多个数据库的查寻与检索(参见本章第二节)。

第二节 核酸序列数据库

核酸序列是了解生物体结构、功能、发育和进化的出发点。国际上权威的三大核酸序列数据库是美国国家生物技术信息中心(NCBI)的 GenBank、欧洲生物信息研究所(EBI)的 ENA、日本遗传研究所(NIG)的 DDBJ,它们收录了所有已知的核酸序列及相关的文献和生物学注释。它们相互合作,建立国际核酸序列数据库协作体(INSDC),每天交换数据,以保持最新、最全的数据,但在数据格式上稍有差别。

一、GenBank

(一) 概述

GenBank (<http://www.ncbi.nlm.nih.gov/genbank>)是由美国国家生物技术信息中心(NCBI)管理和维护的综合性公共核酸序列数据库,收录了所有已知的核酸序列和蛋白质序列,以及相关的文献和生物学注释。数据来源于科研人员提交的序列数据、测序中心提交的大量 EST 序列和其他测序数据以及与其他数据机构协作交换的数据。到 2014 年 8 月,GenBank (203 版)收录的序列数量达到 174 108 750 条,碱基数达 165 722 980 375 个,涉及 280 000 个物种,且数量仍在不断增长。GenBank 的数据可以从 NCBI 的 FTP 服务器上免费下载。NCBI 还提供广泛的数据查询、序列相似性搜索以及其他分析服务。

每条 GenBank 记录包含对序列的简要描述、科学命名、物种分类名称、参考文献、序列特征表以及序列本身。序列特征表包含序列生物学特征注释,如编码区、转录单元、重复区域、突变位点或修饰位点等。

GenBank 的序列记录划分成 12 分类学分区、5 个高通量分区、1 个专利序列分区(PAT)、1 个

全基因组序列分区 (WGS) 和 1 个转录组鸟枪法组装序列分区 (TSA)。其中 12 个分类学分区是细菌 (BCT)、环境抽样序列 (environmental sample sequence, ENV)、无脊椎动物 (INV)、噬菌体 (PHG)、病毒 (VRL)、植物序列 (PLN)、灵长类 (PRI)、啮齿动物 (ROD)、未注释序列 (UNA)、合成序列 (SYN)、其他哺乳动物 (MAM)、其他脊椎动物 (VRT)。5 个高通量分区是 EST、STS、基因组搜索序列 (GSS)、高通量基因组序列 (HTG)、高通量 cDNA (High-throughput cDNA, HTC)。

(二) GenBank 数据的获取

1. Entrez 检索 是 NCBI 生物信息数据库集成检索系统, 可以获取以下 6 类生物信息数据。

(1) 文献及其相关信息: 包括图书、医学主题词、美国国家图书馆馆藏目录 (NLM Catalog)、PubMed 以及 PubMed Central。

(2) 健康相关信息: 包括临床意义的人类变异数据 (ClinVar)、基因型 / 表型交互研究数据 (dbGaP)、遗传检查登记数据 (GTR)、医学遗传学文献和链接 (MedGen)、OMIM 和 PubMed Health。

(3) 基因组相关信息: 包括基因组组装信息 (Assembly)、生物计划提供给 NCBI 的数据 (BioProject)、生物来源材料说明 (BioSample)、基因组或 cDNA 克隆 (Clone)、基因组结构变异研究 (dbVar)、表观基因组数据 (Epigenomics)、生物体基因组测序计划 (Genome)、基因组搜索序列 (GSS)、DNA 和 RNA 序列 (Nucleotide)、探针和引物序列 (Probe)、单核苷酸多态性数据 (SNP)、高能通量 DNA 和 RNA 序列读取数据库 (SRA)、生物学分类数据 (Taxonomy)。

(4) 基因相关信息: 包括表达序列标签 (EST)、基因 (Gene)、功能基因数据 (GEO DataSets)、基因表达和分子丰度分布 (GEO Profiles)、同源基因数据 (HomoloGene)、系统发育和人群研究的序列集 (PopSet)、表达转录集 (UniGene)。

(5) 蛋白质相关信息: 包括保守蛋白质结构域 (Conserved Domains)、蛋白质序列 (Protein)、蛋白质族 (Protein Clusters)、蛋白质结构 (Structure)。

(6) 化学相关信息: 包括与基因、蛋白质和化学相关的分子通路 (BioSystems)、生物活性检测数据 (PubChem BioAssay)、化学成分 (PubChem Compound)、化学物质 (PubChem Substance)。

上述信息, Entrez 可以一次性完成检索, 返回各种信息的检索结果。然后点击某一种信息的结果数字, 即可返回该信息的结果列表, 进一步可获取某一条记录的详细信息。Entrez 可以选择某一数据库单独检索。在结果列表页, 还提供了高级检索, 以便检准信息。

例如查找 H1N1 流感病毒 (H1N1 Flu Virus) 的核酸序列。其检索步骤是:

(1) 进入 NCBI 主页 (<http://www.ncbi.nlm.nih.gov>)。

(2) 在提问框输入 H1N1 Flu Virus。点击“Search”, 得到各个数据库的检索结果 (图 9-2)。

(3) 点击“Nucleotide 63547.....”, 得到 63 547 条核酸序列数据的简要格式 (summary)。

(4) 点击记录的标题, 即可获取该记录的详细信息。

2. BLAST 序列相似性搜索 是 GenBank 数据注释最基础和使用最多的方法。NCBI 提供了一系列的 BLAST 程序, 用于检测查询序列与数据库中的序列的相似性。用户既可以通过 NCBI 网站的 BLAST 主页提交序列进行 BLAST 查询, 也可以通过 NCBI 的 FTP 下载单机版 BLAST 软件, 在本地进行 BLAST 相似性比对。

3. 通过 GenBank 的 FTP 站点 NCBI 以文本文件格式发布 GenBank 数据, 并且以 ASN.1 格式进行内部维护。每两个月的 GenBank 以及 ENA 和 DDBJ 更新的序列数据都可以从 NCBI 的匿名 FTP 服务器上下载 (<ftp://ftp.ncbi.nlm.nih.gov/genbank>), 还可以从 NCBI 的 FTP 服务器免费下载完整的数据库和软件工具。



Search NCBI databases

Help

H1N1 Flu Virus

Search

About 176,794 search results for "H1N1 Flu Virus"

Literature

Books	266	books and reports
MeSH	0	ontology used for PubMed indexing
NLM Catalog	86	books, journals and more in the NLM Collections
PubMed	11,134	scientific & medical abstracts/citations
PubMed Central	10,342	full-text journal articles

Health

ClinVar	1	human variations of clinical significance
dbGaP	0	genotype/phenotype interaction studies
GTR	56	genetic testing registry
MedGen	1	medical genetics literature and links
OMIM	0	online mendelian inheritance in man
PubMed Health	25	clinical effectiveness, disease and drug reports

Genomes

Assembly	0	genomic assembly information
BioProject	4	biological projects providing data to NCBI
BioSample	19	descriptions of biological source materials
Clone	0	genomic and cDNA clones
dbVar	16	genome structural variation studies
Epigenomics	0	epigenomic studies and display tools
Genome	1	genome sequencing projects by organism
GSS	0	genome survey sequences
Nucleotide	63,547	DNA and RNA sequences
Probe	0	sequence-based probes and primers
SNP	0	short genetic variations
SRA	8	high-throughput DNA and RNA sequence read archive
Taxonomy	0	taxonomic classification and nomenclature catalog

Genes

EST	0	expressed sequence tag sequences
Gene	11	collected information about gene loci
GEO DataSets	1,685	functional genomics studies
GEO Profiles	1,498	gene expression and molecular abundance profiles
HomoloGene	0	homologous gene sets for selected organisms
PopSet	216	sequence sets from phylogenetic and population studies
UniGene	0	clusters of expressed transcripts

Proteins

Conserved Domains	0	conserved protein domains
Protein	87,588	protein sequences
Protein Clusters	0	sequence similarity-based protein clusters
Structure	73	experimentally-determined biomolecular structures

Chemicals

BioSystems	67	molecular pathways with links to genes, proteins and chemicals
PubChem BioAssay	132	bioactivity screening studies
PubChem Compound	0	chemical information with structures, information and links
PubChem Substance	7	deposited substance and chemical information

图 9-2 Entrez 对各个数据库的检索结果

二、ENA

(一) 概述

ENA (<http://www.ebi.ac.uk/ena>) 是国际三大核酸序列数据库之一, 创建于 1982 年。现由欧洲生物信息学研究所 (EBI) 管理和维护, 主要收集欧洲产生的核酸序列数据。ENA 整合了原始的序列数据、组装信息和功能注释。到 2014 年 6 月, ENA (120 版) 的核酸序列达到 441 582 575 条, 碱基数达到 883 726 016 221 个。每一条核酸序列记录包括概览 (Overview)、来源特征 (Source Feature(s))、其他特征 (Other Features)、组装 (Assembly)、参考文献、序列 (Sequence) 六个部分。

(二) ENA 数据的获取

1. ENA 检索和浏览 在 ENA 主页 (<http://www.ebi.ac.uk/ena/>, 图 9-3) 提供了文本检索 (Text

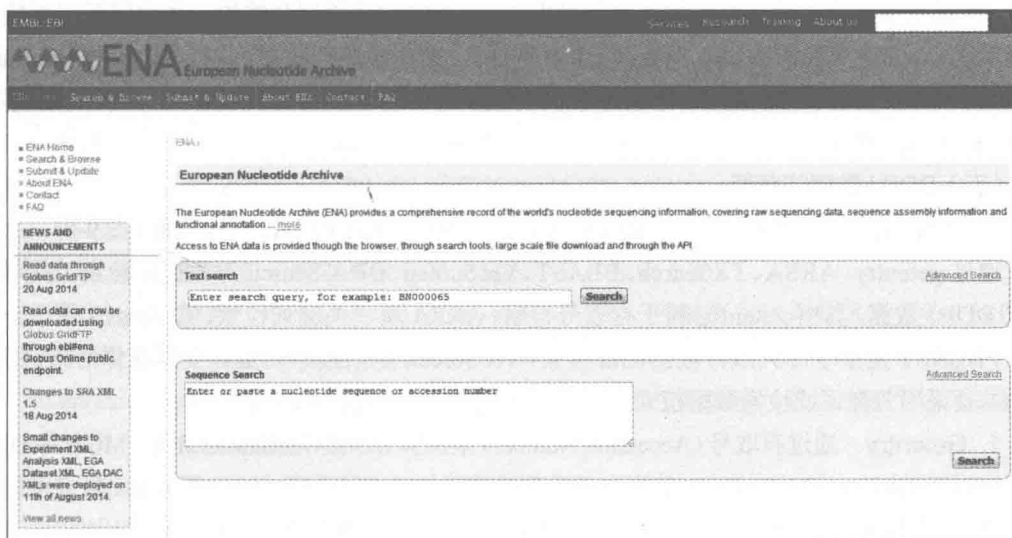


图 9-3 ENA 主页

search)和序列检索(Sequence Search)。如果要进行更多限定,可点击右侧的“Advanced Search”,进入高级检索页面。

2. EBI 检索 ENA 数据 在 EBI 检索服务主页 (<http://www.ebi.ac.uk/ebisearch/>)(图 9-4)提供了基本检索和高级检索以及单库检索功能。

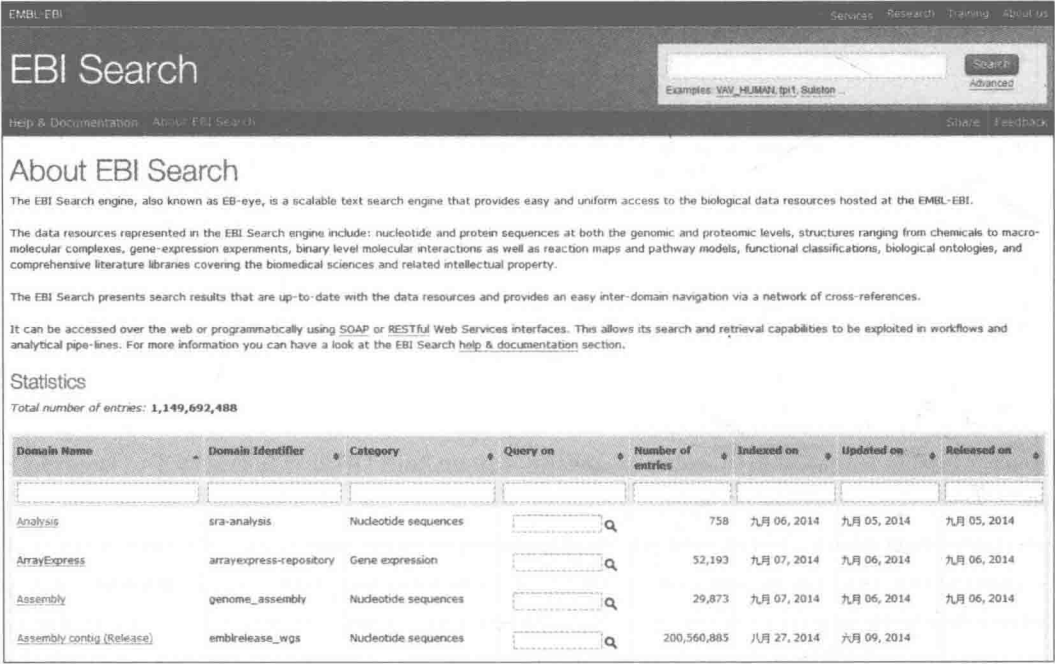


图 9-4 EBI 检索服务主页

3. 通过 ENA 的 FTP 站点下载最新版的 ENA 数据库 通过 ENA 的匿名 FTP 服务器 (<ftp://ftp.ebi.ac.uk/pub/databases/embl/release/>) 下载最新更新的 ENA 数据库。

三、DDBJ

(一) 概述

日本核酸数据库 (DNA Data Bank of Japan, DDBJ, <http://www.ddbj.nig.ac.jp>) 创建于 1986 年, 现由日本国家遗传学研究所的生物信息中心 (CIB/DDBJ) 管理和维护。作为国际三大核酸数据库之一, 它主要负责收集亚洲地区 (主要是日本) 的核酸序列数据, 并与 NCBI 的 GenBank 和 EBI 的 ENA 协作, 同步更新。截至 2014 年 6 月, DDBJ (97 版) 共收录 172 402 324 条序列 161 078 598 329 个碱基。

(二) DDBJ 数据的获取

在 DDBJ 主页的服务栏, 点击“Search/Analysis”, 进入 DDBJ 检索 / 分析页面 (图 9-5)。

通过 getentry、ARSA、TXSearch、BLAST、VecScreen、DRA Search 等数据库检索工具可以获取 DDBJ 数据, 其中 getentry 属于存取号检索, ARSA 属于关键词检索, TXSearch 属于分类检索, BLAST 是基于序列或片段的同源搜索, VecScreen 是扫描序列是否受到载体污染, DAR Search 是采用关键词或检索数据检索元数据。下面主要介绍前 3 种检索工具。

1. Getentry 通过存取号 (Accession Number) 来检索 DDBJ/GenBank/EMBL, MGA 等 DNA 数据库以及 UniPort, PDB, DAD 等蛋白质数据库以及 Patent 专利数据库 (图 9-6), 最多可同时输入 10 个存取号进行检索, 各存取号之间用空格或“,”加以分隔, 连续存取号可用“-”表示多个连续的号码, 例如 D11111, D11112, D11113, D11114 或 D11111-D11114 或 D11111-4。

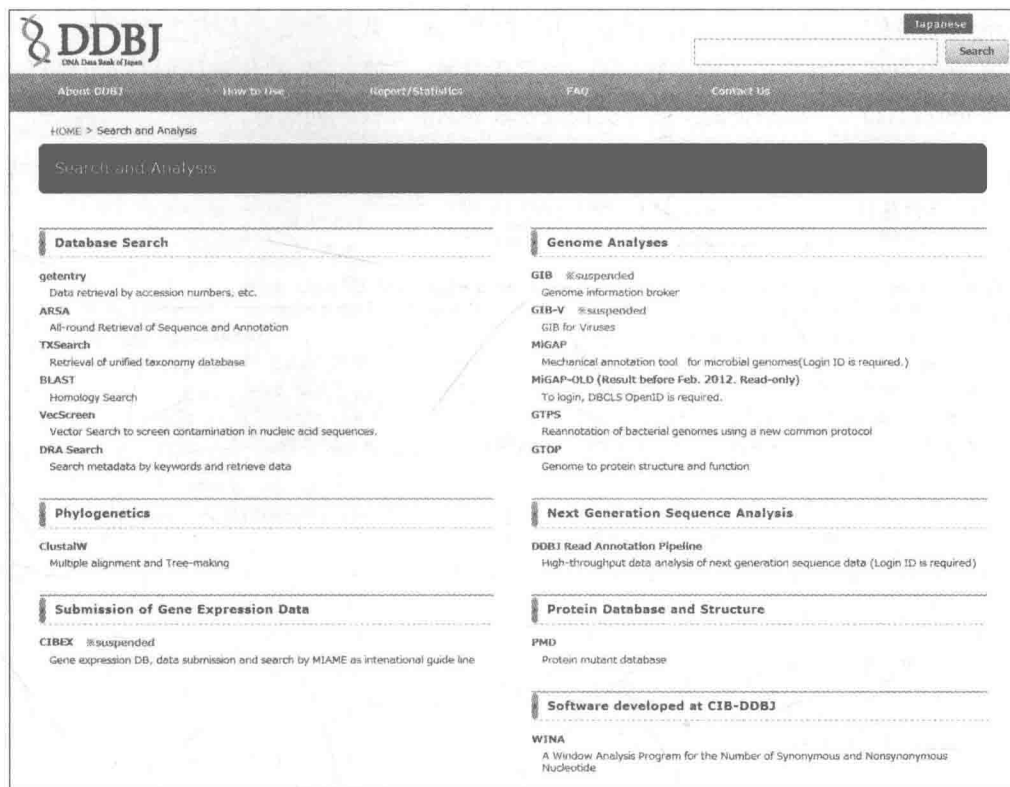


图 9-5 DDBJ 检索 / 分析页面



图 9-6 Getentry 检索页面

通过专利号 (Patent Retrieval Number) 检索时, 必须删去中间的空格, 可省略其中的非字母数字式符号, 如“-”、“-”、“:”和“/”, 例如 JP2006516190-A/72 (专利号), 检索输入格式为: JP2006516190-A72 或 JP2006516190A72 或 JP2006516190-A/72。

2. ARSA 提供了 Quick Search、Advanced Search 2 种检索功能, 支持 AND、OR、NOT 布尔逻辑检索、截词检索、短语检索和限定字段检索等多种检索方法。

3. TXSearch 检索 TXSearch (taxonomy retrieval) 是一种根据生物分类学的分类名称作为关键词进行检索的检索途径 (图 9-7)。

Taxonomic name or ID number 右侧为检索词输入框, 可在此处输入多达 128 个字母的检索词, 并且可用 “*” 作为通配符。

Class of taxonomic name 右侧的下拉菜单可选择命名种类, 有 “scientific name”、“common name”、“acronym”、“synonym”、“anamorph” 和 “All” 六种, 可任选其一进行检索。

Taxonomic rank 用于选择该词的等级, 包括 “All”、“norank” 等 30 个等级词。设置每一页显

示的记录数及结果显示格式后, 点击“Search”, 得到检索结果。

检索结果提供所检索词的详细信息, 如家系名称、其他名称、遗传密码等; 同时还可点击“Go to lower taxa”, 了解该词的下级分类词或获取该词的相关序列信息。

The screenshot shows the TXSearch interface with the following elements:

- Header:** DDBJ DNA Data Bank of Japan, TXSearch Taxonomy Retrieval, [Version 2.08] Updated: Apr. 11, 2011, and a HELP button.
- Instructions:** "When you select [exact matching], you can use * as a wild card in your query. When you search Taxonomy name containing *, in order to distinguish * from a wild card, please input %*."
- Search Fields:**
 - Taxonomic name or ID number: Virus*
 - Class of taxonomic name: scientific name
 - Taxonomic rank: All
 - Max. number of result: 10
 - Would you like to see name(s) in the result as scientific name? Yes (selected) / No
- Buttons:** exact matching, Search
- Results Section:**
 - Result of preliminary search
 - Following information match(es) with your query.
 - Total match: 17 (PAGR 1/2) Next
 - 1. Taxonomic name: Viruses [superkingdom, scientific name]
 - Lineage: [Go to lower taxa]
 - 2. Taxonomic name: Virus-like particle CAK1 [species, scientific name]
 - Lineage: Viruses, unclassified phages [Go to lower taxa]

图 9-7 TXSearch 检索页面及其检索结果

第三节 蛋白质数据库

一、蛋白质序列数据库

(一) PIR

蛋白质信息资源 (the Protein Information Resource, PIR, <http://pir.georgetown.edu/>) 是由美国国家生物医学研究基金会 (National Biomedical Research Foundation, NBRF) 于 1984 年创建的一个综合性公共生物信息学资源, 其目的是支持基因组、蛋白质组和系统生物学的研究, 帮助研究人员鉴别和解释蛋白质序列信息, 研究分子进化、功能基因组和生物信息学分析。

1. PIR 的资源 在 PIR 主页 (图 9-8) 的资源栏目提供以下 6 种蛋白质信息资源。

(1) 集成蛋白质分类数据库 (iProClass): 提供 UniProtKB 和 UniParc 数据库中 NCBI Entrez 特有的蛋白质序列, 与 175 个生物学数据库建立超链接, 主要包括蛋白质家族、功能和通路, 相互作用, 结构和结构分类, 基因和基因组, 本体、相关文献和生物学分类。iProClass 采用数据仓库和超文本导航方法将不同类型的数据集成在一起, 综合揭示蛋白质特征。

(2) 蛋白质本体 (protein ontology, PRO): 提供蛋白质相关记录的本体描述。每一 PRO 词描述了记录的分类, 包括特定修饰形式, 直系同源亚型和蛋白质复合物。

(3) 集成蛋白质文献信息知识库 (integrated protein literature, INformation and Knowledge, iProLINK): 有助于基于文献的数据库文本挖掘, 即条目识别和蛋白质本体开发。

(4) 蛋白质超家族资源 (PIRSF) 采用等级体系的方法将 UniProtKB 序列聚成无重叠的类, 以揭示其进化关系。PIRSF 分类体系是基于全蛋白质而不仅仅是结构域。

(5) PIR 蛋白质序列数据库 (PIR-PSD) 是世界上第一个蛋白质序列分类和注释数据库, 起源于 Dayhoff 编辑出版的 *Atlas of Protein Sequence and Structure* (1965—1978) (蛋白质序列和结构图集 (1965—1978)) 一书, 后来由美国 NBRF、日本国际蛋白质信息数据库 (Japan International



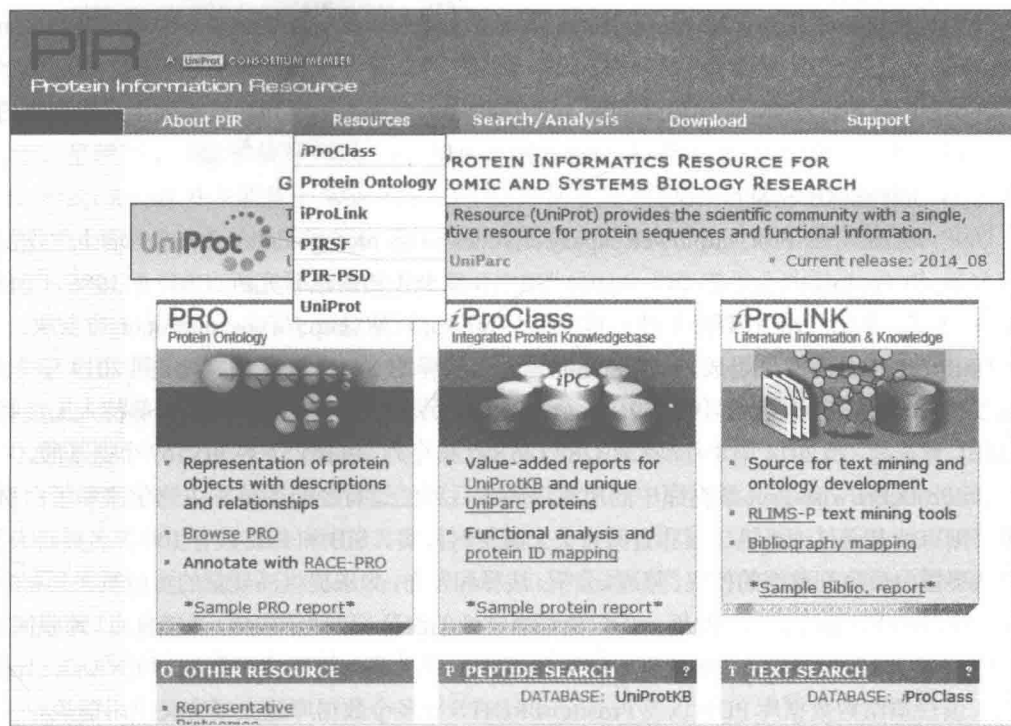


图 9-8 PIR 主页

Protein Information Database, JIPID)与德国慕尼黑蛋白质序列信息中心(Munich Information Center for Protein Sequences, MIPS)合作成立国际蛋白质序列信息中心(PIR-International),共同收集和维护。2002年,PIR联合EBI、SIB成立UniProt联盟,PIR-PSD序列和注释集成到UniProt知识库。UniProt(UniProt Knowledgebase和(或)UniParc)与PIR-PSD建立双向链接和交叉参考,允许用户跟踪以前的IR-PSD记录。2004年12月31日发布了PIR-PSD最终版80.00,以后不再更新。

(6) 通用蛋白质资源(the Universal Protein Resource, UniProt, <http://www.uniprot.org/>): 是一个综合性的蛋白质序列和注释数据库。主要由3个部分组成: ① UniProt Knowledgebase (UniProtKB): 这是蛋白质序列、功能、分类、交叉引用等蛋白质知识库,记录经过人工筛选和注释; ② UniRef(UniProt Non-redundant Reference)数据库: 将密切相关的蛋白质序列组合到一条记录中,以便提高搜索速度,目前,根据序列相似程度形成3个子库,即UniRef100、UniRef90和UniRef50; ③ UniParc(UniProt Archive): 是UniProt存档库,收录所有蛋白质序列。用户可以通过文本查询数据库,可以利用BLAST程序搜索数据库,也可以直接通过FTP下载数据。

2. PIR的检索与分析工具 在PIR主页的检索/分析(Search/Analysis)栏目,提供了多种检索与分析工具。

(1) 文本检索(Text Search): 输入文本词或标识号检索iProClass的单个蛋白质,或者PIRSF数据库中的蛋白质家族,支持限定字段检索、布尔逻辑检索。

(2) 批量检索(Batch Retrieval): 采用特定的标识号或者组合不同的标识号从iProClass检索多个条目,或者从PIRSF数据库检索多个蛋白质家族。

(3) 肽匹配(Peptide Match): 精确查找肽序列(3~30个氨基酸残基)的匹配序列。2个数据库可以使用该检索,一是UniProtKB,包括蛋白质功能信息、审核和注释信息;二是UniRef100,包括已知序列和片段。

(4) 多重序列比对(Multiple Alignment): 输入多个FASTA格式的序列或多个UniProtKB标识号,得到CLUSTALW多重比对的结果,以及邻接树和PIR交互树和比对浏览。

(5) 两序列比对 (Pairwise Alignment): 输入两个序列或者 UniProtKB 标识号, 得到 Search Smith-Waterman 全长比对结果。

(6) 分子量和组成 (Molecular Weight and Composition): 基于序列的氨基酸组成检索蛋白质的分子量。

(二) UniProtKB/Swiss-Prot

UniProtKB/Swiss-Prot (http://web.expasy.org/docs/swiss-prot_guideline.html) 由瑞士生物信息学研究所 (Swiss Institute of Bioinformatics, SIB) 和欧洲生物信息研究所 (EBI) 于 1986 年创建, 自 2002 年起, 由 UniProt 联盟维护和管理, 通过 UniProt 网站 (<http://www.uniprot.org/>) 获取。

UniProtKB 由两部分组成, 一部分是经过人工注释的 Swiss-Prot 数据库, 到 2014 年 7 月, 收录了 546 000 条序列, 包含 194 259 968 个氨基酸。另一部分是计算机分析的等待人工注释的 TrEMBL 数据库, 到 2014 年 8 月, 收录了 82 126 897 条序列, 包含 25 939 595 569 个氨基酸。

UniProtKB/Swiss-Prot 数据库中的所有序列条目都经过有经验的分子生物学家和蛋白质化学家仔细审读并通过计算机工具和查阅有关文献资料核实。SIB 和 EBI 共有 100 多名科研人员, 专门从事蛋白质序列数据的搜集、整理、分析、注释和发布, 力求提供高质量的蛋白质序列和注释信息。UniProtKB/Swiss-Prot 的每一条记录都有详细的注释, 包括结构域、功能位点、跨膜区域、二硫键位置、翻译后修饰、突变体等。该数据库中的记录还与核酸序列数据库 ENA/GenBank/DBJ、蛋白质结构数据库 PDB 以及 Prosite、PRINTS 等多个数据库建立了交叉引用链接。

(三) OWL

OWL (<http://www.bioinf.manchester.ac.uk/dbbrowser/OWL/index.php>) 是 1994 年由英国里兹 (Leeds) 大学和 Warrington 的 Daresbury 国家实验室合作创建并维护的一个复合型数据库, 其数据来源于 SWISS-PROT、PIR (1-3)、GenBank (翻译的)、NRL-3D 等 4 个公共获取的原始数据库, 去重后整合而成的非冗余蛋白质序列数据库。

OWL 数据库在构建过程中, 考虑到每个数据库所包含序列信息的情况, 赋予它们不同的优先级, SWISS-PROT 数据库的优先级最高。在对数据的处理上, 不仅删除与某一序列完全相同的序列条目, 也剔除与某一序列相差个别氨基酸残基的序列条目。因此, OWL 数据库是一个具有较小冗余度的蛋白质序列数据库。尽管如此, OWL 数据库也会有一些错误, 即来自一次数据库的错误序列, 此外, OWL 数据库更新较慢。

OWL 提供了存取号、数据库编码、文本词、序列、题名、作者、查询语种等多种检索方式。

二、蛋白质结构数据库

蛋白质结构数据库则是非常重要的生物信息学资源。20 世纪 90 年代以来, 越来越多的蛋白质结构被测定, 蛋白质结构分类的研究不断深入, 出现了蛋白质结构、蛋白质家族、折叠、结构域、回环等多种类型的结构数据库。

1. PDB (Protein Data Bank, <http://www.rcsb.org/pdb/>) 创建于 1971 年, 是国际上最著名、最完整的蛋白质三维结构数据库。最先由美国 Brookhaven 国家实验室负责维护和管理, 从 1998 年开始, 由结构生物信息学合作研究协会 (RCSB) 负责管理。与 GenBank 核酸序列数据库一样, 科研人员可以通过网络直接向 PDB 数据库提交数据。

PDB 是目前最主要的蛋白质分子结构数据库。随着晶体衍射技术的不断改进, 结构测定的速度和精度也逐步提高。20 世纪 90 年代以来, 多维磁共振成像技术的成熟, 使得那些难以结晶的蛋白质分子的结构测定成为可能。到 2014 年 9 月 2 日统计, PDB 数据库已经收录了利用 X 线衍射、NMR、电子显微镜实验数据或理论计算得出的蛋白质、核酸、蛋白质 / 核酸复合物等结构数据 103 015 条, 而且数据增长速度相当快。这些数据由 RCSB 组织专业人员注释并向全世界免费公布。



PDB 数据库以文本文件格式存放数据,每条记录即是一个独立的文件,包括物种来源、化合物名称、原子坐标、结构提交者以及有关文献等基本注释信息。此外,还包括分辨率、结构因子、温度系数、主链数目、配体分子式、金属离子、二级结构信息、二硫键位置等和结构有关的数据。虽然用文字编辑软件可以查看 PDB 数据库的记录,但是无法直观地反映分子的空间结构。因此,RCSB 开发的基于 Web 的 PDB 数据库摘要显示系统,只显示主要信息。如需进一步了解其详细信息,或查询其他蛋白质结构信息资源,可以点击该页面左栏中的相应图标,并将 RasMol、Cn3D 等分子结构软件集成在一起,具有数据分析和图形显示功能。

(1) PDB 检索:提供了 2 种检索方式:①基本检索,通过所有字段(Everything)、作者(Author)、大分子名称(Macromolecule Name)、序列(Sequence)和配体(Ligands)进行检索;②高级检索(Advanced Search),可以选择某一字段(如结构注释、结构特征等)、实验方法(如 X 射线、NMR 等)和分子类型(如蛋白质、核酸等)进行检索,支持 AND 逻辑检索。

(2) PDB 浏览:提供了 13 种浏览方式:①解剖治疗化学分类(Anatomical Therapeutic Chemical, ATC),是 WHO 用于药物的一种分类体系,分为 14 个大类;②生物学过程(Bio.Process);③ CATH 分类(CATH);④细胞成分(Cellular Component);⑤酶分类号(EC Numbers);⑥基因组定位(Genome Location);⑦医学主题词(MeSH);⑧膜蛋白(Membrane Prot.);⑨分子功能(Molecular Function);⑩蛋白质对称性(Protein Symmetry);⑪ SCOP 分类(SCOP);⑫来源生物体(Source Organism);⑬转运蛋白分类(Transporter Classification)。

2. NCBI 结构数据库(<http://www.ncbi.nlm.nih.gov/Structure/index.shtml>) 由 NCBI 计算生物学部结构组开发,主要包括 4 个方面的数据。

(1) 大分子结构(Macromolecular structures):包含丰富的生物学功能和进化关系等信息,不仅可以获取其结构数据,还可以获取与之相关的文献、蛋白质和核酸序列、化学性质、生物分子相互作用等信息。此外,通过“Related Structures”获取三维结构同源的蛋白质。通过蛋白质查询序列和蛋白质结构序列的比对,可以浏览相应的三维结构和序列结构关系和活性位点。

(2) 保守结构域和蛋白质分类(Conserved domains and protein classification):保守结构域数据库(the Conserved Domain Database, CDD)收集了多重比对揭示的保守结构域,以及 NCBI 采用三维结构信息定义的功能区域,揭示了序列/结构/功能三者之间的关系。

(3) 小分子及其生物学活性(Small molecules and their biological activity):PubChem 项目是 NIH 分子库路线图计划的一个重要组成部分,提供了小分子的生物活性信息。PubChem 由三个数据库组成:PCSubstance,PCBioAssay and PCCompound,前两个是科研人员发送的化学物质,包括医药及其生物活性;第三个是衍生的、非冗余的蛋白质成分数据库。

(4) 生物系统(Biological Systems):主要收集具有直接或间接相互作用的生物分子。

结构数据库是 Entrez 系统的一部分,利用 Entrez 对结构数据库进行文本查询。结构数据库检索项包含 PDB、MMDB 的存取号,源自 PDB 注释记录的自由文本,作者名及其他书目检索项。

VAST(矢量分析检索工具)用于检索三维结构的相似性。它舍弃了序列信息,而单纯依靠由二级结构衍生的三维矢量指标进行检索。VAST 具有检索结构相似而序列不相似的生物高聚体的能力。如同 BLAST 一样,VAST 可以 N*N 方式操作数据库中的所有条目,并存储结果,以利于 Entrez 页面快速检索。利用 VAST 算法,对 10 000 多个局部结构一一匹配比较,并记录了结构之间的匹配与重叠关系。

Cn3D 是一种新的三维结构浏览器,用于浏览 MMDB 数据记录。由于 MMDB 数据记录克服了一些 PDB 条目数据化学图像模棱两可的缺点,并且所有的键信息是“显性的”,所以 Cn3D 具有可靠地显示三维数据库结构的能力。

Cn3D 提供了一套“控制面板”,可以在三维图像的一侧隐藏或显现。例如在三维图像顶端显现的 Viewer Control 面板,用于逐个显示 NMR 获得的多个结构“集合总体”或 VAST 相似性关系

图像中的重叠结构。“GO”键使图像形成动画,当图像以动画形式显示时。用户可以旋转或放缩结构。Cn3D4.3.1 将具有完整的状态存贮能力,使结构彩色化描述,以 ASN.1 结构记录形式存贮信息和手工编制 PDB 文件或手写文本更新成为可能。这些信息可以被不同工作平台的用户共享。

3. SCOP (Structural Classification of Proteins, <http://scop.mrc-lmb.cam.ac.uk/scop/>) 是英国医学研究委员会分子生物学实验室和蛋白质工程中心开发的,基于 Web 的蛋白质结构分类数据库。SCOP 的目标是提供关于已知结构蛋白质之间的结构和进化关系的信息,所涉及的蛋白质包括结构数据库 PDB 中的所有条目。SCOP 数据库除了提供蛋白质结构和进化关系信息外,对于每一个蛋白质还包括下述信息: PDB 的链接、序列、参考文献、结构的图像等。从目前的技术来看,很难借助于自动的序列和结构比较工具发现蛋白质之间的结构和进化关系,因此, SCOP 的结构分类主要是通过人工来完成的,通过图形显示器观察和比较蛋白质结构,并借助于软件工具进行分析,如同源序列搜索工具。

SCOP 按结构和进化关系对蛋白质分类,主要划分为全 α 型 (All alpha proteins)、全 β 型 (All beta proteins)、以平行折叠为主的 α/β 型、以反平行折叠为主的 $\alpha+\beta$ 型等 11 类。然后,再将属于同一结构类型的蛋白质按照折叠 (Fold)、超家族 (Superfamily)、家族 (Families)、蛋白质结构域 (Protein domains)、单个 PDB 蛋白质结构层次组织起来。

属于同一个折叠类的蛋白质具有相似的折叠结构。如果两个蛋白质具有相同的主要二级结构,并具有相同的拓扑连接,那么,这些蛋白质就具有共同的折叠。具有相同折叠的不同蛋白质,通常有不同大小和不同构象的外周二级结构元素及转向区域。在某些情况下,这些不同的外周区域可能构成完整结构的一半。按照相同折叠放入同一个折叠分类中的蛋白质可能没有共同的进化源,结构相似性可能缘由蛋白质倾向于形成一定堆积和一定拓扑结构的物理和化学特性。

具有明显进化关系的蛋白质聚集到一个家族中,这意味着两个蛋白质之间的等同氨基酸残基数超过 30%。在某些情况下,虽然两个蛋白质序列不相似,但它们具有相似的结构和相似的功能,表明属于同一个家族。例如,许多球蛋白虽然序列相同部分只达到 15%,但确实属于同一个家族。

超家族中的成员具有远缘进化关系,具有共同的进化祖先。有些蛋白质,它们序列之间的相似性较低,序列等同部分短,但是结构和功能特征显示可能有一个共同的进化祖先,对于这些蛋白质将它们放入一个超家族中。

2009 年 6 月 SCOP 不再更新,在网上可查到 2009 年 6 月发布的 1.75 版 (图 9-9),包含 11 个类,1195 折叠,1962 个超家族,3902 家族,110800 结构域。可以从分类体系顶层浏览 (top of the hierarchy),也可进行关键词检索。

2009 年 6 月以后,被 SCOP2 (<http://scop2.mrc-lmb.cam.ac.uk/>) 和 SCOPe (<http://scop.berkeley.edu/>) 所继承和发展。

4. CATH (<http://www.cathdb.info/>) 是一个蛋白质结构域分级分类数据库,由英国伦敦大学开发和维护。CATH 主要有四个层次:类 (Class, C)、构架 (Architecture, A)、拓扑 (Topology, T) 和同源超家族 (Homology superfamily, H)。2013 年 3 月 26 日统计显示,CATH4.0 版共收集了 277687 结构域,2738 个超家族。

CATH 的类与 SCOP 数据库类似,主要根据二级结构划分为 α 、 β 、 $\alpha+\beta$ 、 α/β 等结构类型。构架层次描述结构域的形状,主要由其二级结构的取向决定。最终结构的判定是由专家人工完成,而不是由程序自动进行。

CATH 的拓扑层描述的是折叠家族的水平。家族的概念在 SCOP 和 CATH 中有所差别,SCOP 家族的概念强调的是序列的相似性,而 CATH 的家族则强调的是二级结构的相似性,利用 SSAP 算法将蛋白质结构域聚类成不同的家族。属于同一个拓扑的蛋白质并不一定是同源的。与此对应,同源超家族层次则将对具有同源性的蛋白质聚类在一起。



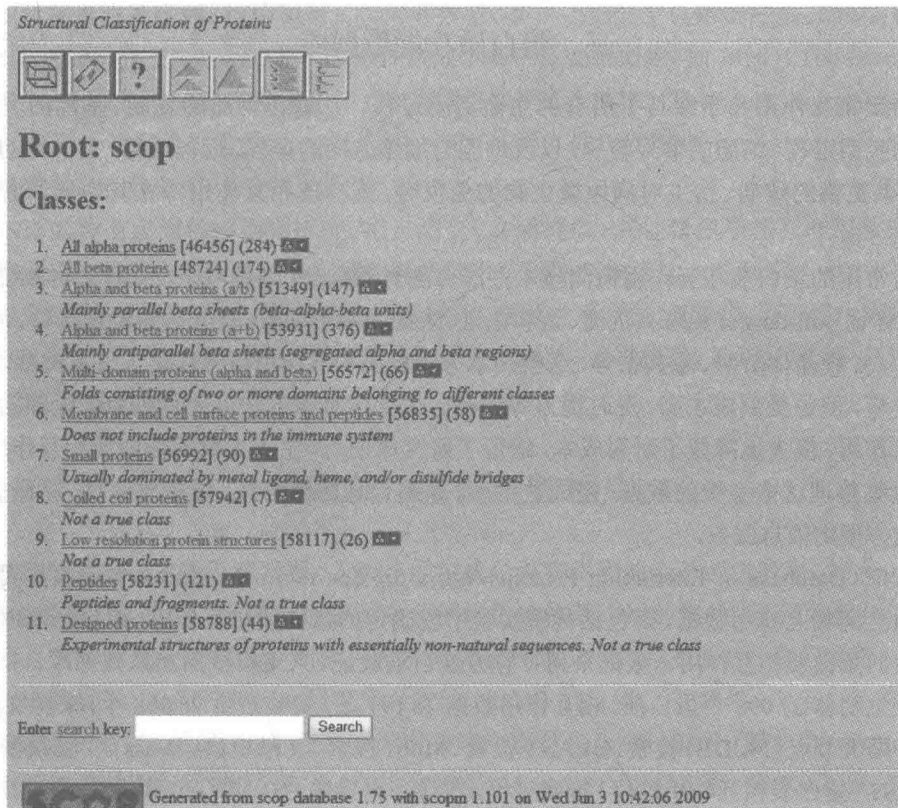


图 9-9 SCOP 的结构

在 CATH 主页可以用 PDB 存取号、CATH 存取号或关键词进行搜索。如输入 1PBO，得到 2 个结构域、2 条链以及 PDB 描述(图 9-10)。点击其中的超链，可以获取更详细的信息。

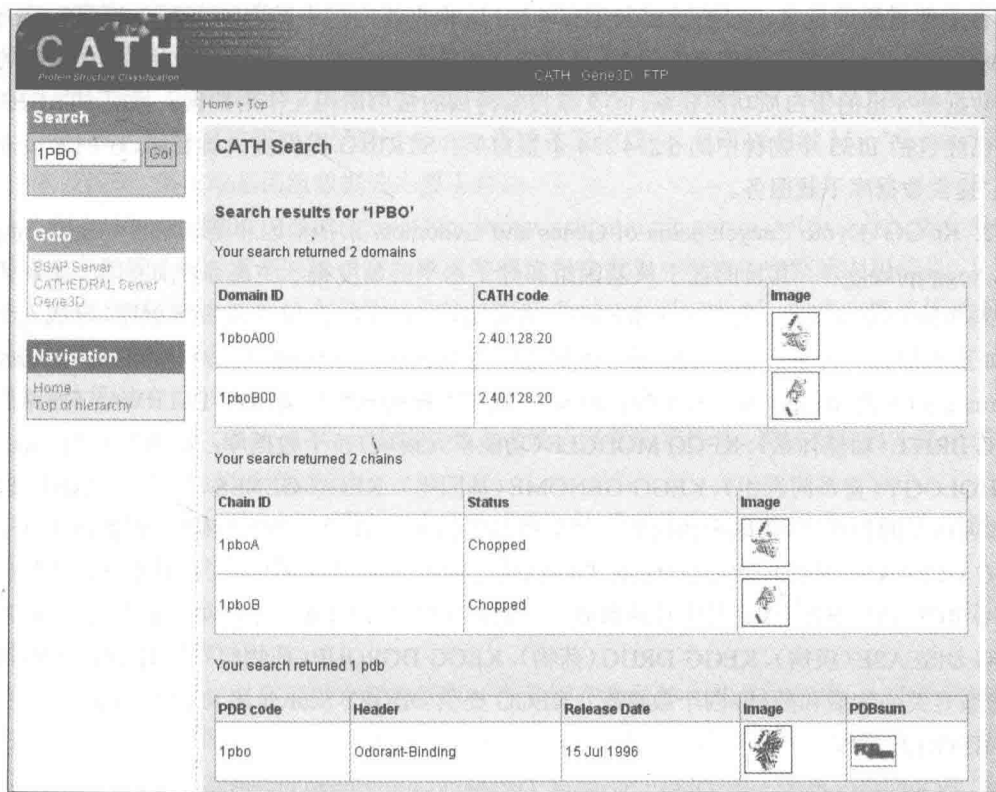


图 9-10 CATH 数据库检索结果

三、蛋白质功能数据库

蛋白质相互作用几乎参与了所有的生命活动过程。从遗传物质的复制、基因的表达调控到细胞的代谢过程、细胞的信号转导,以及细胞与细胞之间的短程、远程通信,生物体的形态形成、病原微生物的致病、宿主对病原微生物的免疫等,蛋白质相互作用都在其中扮演着重要的角色。

目前常用的蛋白质相互作用检测技术方法有酵母双杂交、细菌双杂交、哺乳动物细胞双杂交、GST 标记 pull-down、免疫共沉淀、亲和色谱、表面等离子共振、荧光共振能量转移、质谱、蛋白质芯片、X 线晶体衍射、磁共振等,这些技术为蛋白质相互作用研究作出了重大贡献,积累了宝贵的资料,但是受到成本高、费时费力等限制。生物信息学方法以计算机为手段辅助研究蛋白质相互作用,极大地降低了研究成本,缩短了研究周期,产生了大量的蛋白质相互作用数据。科学家收集整理这些宝贵的数据,建成数据库,并整合基因组信息、蛋白质信息、注释信息等,提供快速方便的信息服务。

1. DIP (Database of Interacting Proteins, <http://dip.doe-mbi.ucla.edu/>) 专门收集实验确定的蛋白质之间相互作用数据,既包括传统实验手段确定的蛋白质相互作用,也包括高通量实验手段确定的蛋白质相互作用。数据采用计算方法自动验证、人工审核后加入数据库。DIP 收录了 693 种生物体 27 098 个蛋白质,相互作用数据 78 191 条,实验数量 76 008 个,来源文献 7092 篇。DIP 提供节点 (NODE) 检索、BLAST 检索、Motif 搜索、文献检索、IMEx (实验数据) 检索、pathBLAST (蛋白质相互作用网络) 搜索等检索服务,也提供 PSI-MI (1.0 和 2.5 版),MITAB2.5 和 XIN 格式的数据下载。

2. STRING (Search Tool for the Retrieval of Interacting Genes/Proteins, <http://string.embl.de/>) 不仅存贮实验确定的蛋白质相互作用数据,而且还存贮预测的蛋白质相互作用数据,并对各种预测方法的准确性给出了相应的权重,对于采用经典实验方法研究蛋白质的功能、生物学意义具有非常重要的意义。它的数据来源有 4 种:①根据基因组中基因的上下文关系 (genomic context) 预测得到的蛋白质相互作用;②高通量实验技术产生的蛋白质相互作用;③由保守的共表达数据推导出的蛋白质功能联系;④文献搜索得到的蛋白质相互作用数据。目前 STRING 数据库已经包括 1133 个物种中的 5 214 234 个蛋白质。STRING 提供蛋白质名称、序列等检索途径,也提供数据库下载服务。

3. KEGG (Kyoto Encyclopedia of Genes and Genomes, 京都基因和基因组百科全书, <http://www.kegg.jp/kegg/>) 其目的在于从基因组和分子水平的角度揭示生态系统如细胞、生物体和生态系统的功能和活动。它是一个数据库集成系统,目前整合了 17 个主要数据库,分成 4 个大类,如系统信息 (systems information)、基因组信息 (genomic information)、化学信息 (chemical information) 和健康信息 (health information)。其中系统信息包括 KEGG PATHWAY (通路图)、KEGG BRITE (功能体系)、KEGG MODULE (功能单元模块) 三个数据库;基因信息包括 KEGG ORTHOLOGY (直系同源组)、KEGG GENOME (基因组)、KEGG GENES (基因)、KEGG SSDB (与基因相关的相似序列) 四个数据库。化学信息包括 KEGG COMPOUND (代谢物及小分子)、KEGG GLYCAN (多糖)、KEGG REACTION (生化反应)、KEGG RPAIR (配对化学转化反应)、KEGG RCLASS (RPAIR 定义的反应级别)、KEGG ENZYME (酶) 六个数据库;健康信息包括 KEGG DISEASE (疾病)、KEGG DRUG (药物)、KEGG DGROUP (药物组) 和 KEGG ENVIRON (与健康有关的物质和药材) 四个数据库。KEGG 提供 DBGET Search 和 RDB Search,可以免费下载 KEGG 数据库。

4. BOND (Biomolecular Object Network Databank, <http://bond.unleashedinformatics.com/>) 是生物分子网络数据库,创建于 2005 年,由 Unleashed Informatics 公司管理和维护,主要收录



生物分子及其蛋白质序列、结构和相互作用等数据,其中生物分子包括蛋白质、DNA、RNA、配基、复合体、基因、光子(photon)等。BOND 集成了多个重要数据库,如 BIND(The Biomolecular Interaction Network Database,生物分子相互作用网络数据库)、SMID(The Small Molecule Interaction Database,小分子相互作用数据库)、Genbank、GO(基因本体)、OMIM、conserved domains(保守功能域)、交叉参考数据库、完整基因组等。

BOND 的数据来源包括期刊提交、用户提交、文献提取、其他蛋白质相互作用数据库导入等。各种来源的数据收集后经过专家依照严格的审核程序审核后加入数据库。因此,不论是从数据的综合性、数据的可靠性,还是查询的多样化、便捷性等方面,BOND 是研究生物分子相互作用的首选数据库。BOND 免费注册,为用户提供了浏览、文本检索、标识符检索、高级特定字段检索、数据下载等服务。

第四节 基因组数据库

随着核酸测序技术的迅速发展,人类已经完成了许多生物的全基因组测序,如人、小鼠、大鼠、大肠埃希菌、酵母等。全基因组数据对于认识生物体基因组的奥秘、了解生物体的生长发育的规律具有非常重要的意义。国际上许多研究机构建立基因组数据库,专门收集、存储、管理和分析这些数据,如英国 Roslin 研究所开发的 ArkDB 收集猪、牛、绵羊、山羊、马、鹿、狗、鸡等基因组数据;美国、英国、日本等基因组中心建立斑马鱼(zebrafish)、罗非鱼(tilapia)、鲑鱼(salmon)等鱼类基因组数据库;英国谷物网络组织(CropNet)建有大麦、玉米、高粱、菜豆等农作物以及紫苜蓿、牧草、玫瑰等基因组数据库。

一、Entrez Genomes

NCBI 提供的 Entrez Genomes (<http://www.ncbi.nlm.nih.gov/genome>) 是世界上最大、最完整的多物种的基因组资源,截至 2014 年 9 月 20 日,已收录 10488 个物种,按生物体分成真核生物(eukaryotae)、原核生物(prokaryotes)、病毒(viruses)和质粒(plasmids)四大类。该数据库提供各种基因组图谱,包括染色体、序列图谱、遗传图谱、物理图谱和连续子(contigs)图谱。

Entrez 基因组数据的获取可以通过 Entrez 进行关键词搜索,也可以通过 Map Viewer 进行浏览、检索及编辑,是获取基因组数据的主要工具。

1. Map Viewer 概况 Map Viewer (<http://www.ncbi.nlm.nih.gov/mapview/>) 是汇集众多基因组图谱和序列信息的图谱浏览器。它允许用户浏览和检索某一生物体完整基因组信息,又允许用户在序列水平,通过浏览单个染色体图谱或某染色体上的特定区域,探查完整基因组信息。对于组装好的完整基因组,Map Viewer 从 4 个层次揭示其信息。

(1) 生物体主页(Home Page): 提供某一生物体基因组概要信息。

(2) 基因组浏览(Genome View): 以图形方式显示完整的染色体,通过相应的关键词进行检索,显示相应染色体的匹配位点。

(3) 图谱浏览(Map View): 给出所选染色体的一种或多种感兴趣的图谱,并与主图谱并列显示,允许用户以不同的分辨率浏览感兴趣的区域。

(4) 序列浏览(Sequence View): 可显示相应染色体特定区域的序列信息,并以图形方式给出相应区域的生物学特征。

通过 Map Viewer,用户可以方便地获取一个基因在染色体上的具体位置、一条染色体上排列的基因及其顺序、染色体特定区域的基因种类、染色体特定区域的碱基顺序、两个基因的距离、染色体两个位点之间的序列及细胞遗传图谱等信息,并可根据两个图谱中的标记进行对比分析,以及在已知某一基因在遗传图谱中的位置,确定其在物理图谱中的位置。

部分染色体具有多种图谱信息,一般包括细胞遗传图谱、连锁图谱、杂交图谱和序列图谱,图谱的数量与类型因不同物种基因组研究进程而异,可根据其共有的标记、基因名称、序列进行对比分析。Map Viewer 谱可同时显示多个图谱,并允许指定一个为主图谱(Master Map),显示时将置于窗口的最右侧,并给出其描述性文本信息。

2. Map Viewer 的使用 在 Map Viewer 主页(<http://www.ncbi.nlm.nih.gov/mapview/>),从 Search 提问框下拉式菜单中选择 Homo sapiens,点击“Go”,或者在主页面结构树中直接点击 Mammals 类 Primates 的学科名列表中 Homos apiens 的 Annotation Release 106,返回 Homo sapiens genome view 页面。用户若查找图谱上的任何信息,如基因名、基因库存取号、标记物名称或疾病名称等,只需在 Search for 提问框中输入相应的检索词,点“Find”即可。

例如,输入年龄相关的黄斑病易感因子(age-related maculopathy susceptibility 2)基因名“ARMS2”,点击“Find”后,系统会返回 78 条记录,在右侧 Quick Filter(快速过滤)区,选择 Gene,得到 4 条检索结果(图 9-11)。图 9-11 上方显示了人类所有染色体图形,并用指针指出 ARMS2 在第 10 号染色体长臂上的位置及其命中的记录数;页面的下方则表明了 ARMS2 所在染色体(Chr)、组装(Assembly)的数据库(如 reference、CHM1_1.1-Primary、HuRef-Primary)及检出记录(Match)、图谱元素(Map element)、数据类型(Type)和图谱(Maps)。点击“Map element”栏相应的选项,如 ARMS2,可以得到该基因图谱的详细信息(图 9-12)。其他图谱可通过“Maps & Options”进行修改。

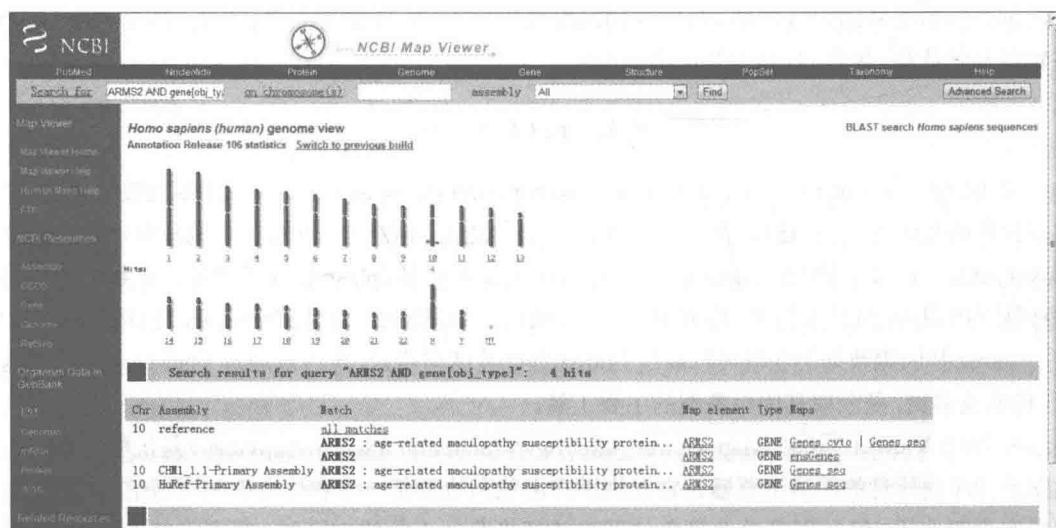


图 9-11 ARMS2 基因检索结果

图 9-12 中的基因序列图谱为主图谱,提供最详细的资料。在 ARMS2 右侧的箭头(粉红色区域内)表明基因转录的方向。点击 ARMS2,进入 GenBank 数据库,可获取该基因的相关信息。此外,与其他数据库建立了许多链接(Links),提供了相关证据(E)、染色体定位(Cyto)和简要描述(Description)。

- (1) OMIM: 人类孟德尔遗传学数据库有关该基因与临床疾病方面的信息。
- (2) HGNC: 人类基因组命名委员会对该基因的命名信息。
- (3) sv: 序列浏览(Sequence Viewer),表明基因在基因组克隆连续子上的位置,包括核酸和编码的蛋白质序列。
- (4) pr: 该基因的蛋白质序列。
- (5) dl: 序列下载(Sequence Download),允许使用者以文体格式重新获取某一区域的基因

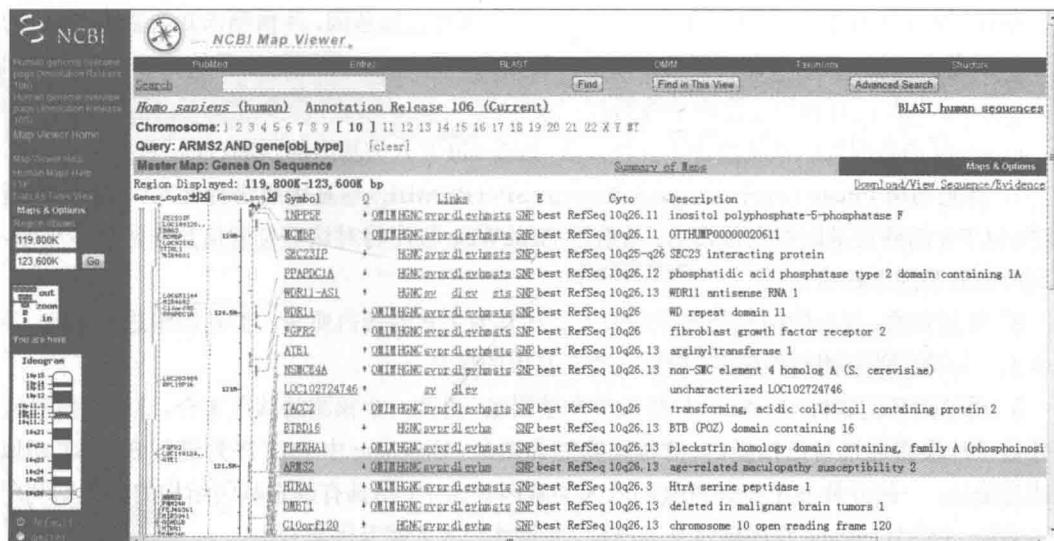


图 9-12 ARMS2 基因图谱信息

组序列,能很容易地替换序列显示的区域。虽然点击 Download/View Sequence/Evidence 超链接也可打开相同的窗口,但二者在显示的起始位点等方面有差别。

(6) ev: 证据浏览 (Evidence Viewer), 能为使用者提供证据浏览, 可显示支持某特定基因模型的生物学证据。这个视图显示了所有的标准序列模型 (RefSeq)、基因库 mRNAs (GenBank mRNAs)、转录子 (无论注解的、已知的或潜在的) 及与基因组 contig 进行序列对齐比较的表达序列标签 (ESTs)。用户通过点击任意的证据浏览页上的 Evidence Viewer Help 链接进入 NCBI 网页, 可进一步获取证据浏览信息。

(7) hm: 指同源基因 (Homologene), 为 NCBI 的人小鼠同源图谱的链接, 显示人类和小鼠之间同源的基因组序列。

(8) sts: 该基因的 STS 序列。

(9) SNP: 该基因的单核苷酸多态性数据。

(10) cyto: 显示该基因在细胞遗传学图谱中的位置, 橙色条带显示基因位置。图 9-2 显示的是 10q26.13 区域。屏幕左侧工具条上的染色体小图形显示该区域在染色体上的具体位置。10q26.13 属细胞遗传学位点。对于 10q26.13 来说, 10 代表染色体号, q 代表染色体的长臂, 26.13 代表染色体上的特定区带, 该值越高, 说明此带离中央着丝粒越远。

(11) Genes seq: 为基因序列图谱, 可用于查看基因结构, 并能显示所有已经被绘制到基因组中的外显子。

(12) 蓝色工具条上的缩放控制区 (Zoom in/out): 可进行图形放大或缩小操作, 缩小一个水平可显示 1/100 的染色体区域。另外, 点击屏幕左边工具条中的 Data As Table View, 可找到包括隐藏在此窗口中一个基于文本的表格的全部资料。

(13) 通过该页面上的其他染色体号, 用户可打开相应的染色体图形。每一染色体图形都会列出基因总数及所显示的区域。

二、Ensembl

Ensembl (<http://www.ensembl.org/index.html>) 是一个有关人类基因组及其他物种基因组的综合数据库, 始建于 1996 年, 由 EMBL-EBI 和 Sanger 研究所共同开发, 包括人类、小鼠、斑马鱼、马、大熊猫等几十种生物基因组的注释分析。

Ensembl 主要根据已经测得的基因组序列, 定位所有已知基因, 并预测未知新基因, 同时为这些基因提供功能、疾病相关特征等方面的注释信息。Ensembl 基因组注释的核心部分是基因预测, 这部分工作的好坏直接影响最终结果。

Ensembl 的基因预测系统整合了多种方法, Ensembl 对人类基因预测过程, 包括:

1. 首先利用 Pmatch 程序将已知人类蛋白 (SP-TREMBL) 与基因组进行比对, 得到这些蛋白对应 CDS 的确定基因组定位信息; 随后用 GeneWise 程序将对这些位置信息作进一步处理, 得到准确的基因结构信息。

2. 将现有的、尚不能确定全部序列的人类和其他生物的蛋白质与人类基因组进行比较, 并整合上一步得到的基因结构信息, 得到人类新基因的集合。

3. 用基因预测软件 GenScan 扫描人类全基因组, 得到一个预测的肽段集合, 这些肽段对应的外显子片段将通过与已知蛋白、脊椎动物的 mRNA、UniGene 中 EST 序列进行 BLAST 相似性比对验证, 一部分外显子序列直接归入人类基因集合中, 不具有确定基因结构的部分则利用 GeneWise、EST2Genome 程序确定基因结构, 并最终归入人类基因集合中。

Ensembl 包括染色体定位、基因组成、EST 拼接、同源性比较等多种信息, 还包括与数据库检索系统 SRS 的链接、BLAST 搜索等功能, 对于这些数据信息, Ensembl 通过不同形式进行展示, 如通过 Map View 显示基因在染色体上定位和分布情况, 通过 Gene View 显示基因转录和基因结构信息; 通过 Protein View 显示预测蛋白质的序列和功能信息等。

Ensembl 的人类基因组主页可以输入关键词 (如 chromosome X 或 12: 10000..200000 或 BRCA2) 进行检索。例如遗传性血红蛋白沉着症 (hereditary hemochromatosis) 基因 “HFE”, 点击 “Go” 后, 系统返回 4055 条记录 (图 9-13)。从左栏对结果分子类型聚类来看, 基因记录 177 条, 转录本 254 条, 变异序列 2485 条; 从物种分类来看, 人类最多, 741 条。可以进一步选定某一类型, 进行限定检索。右栏是检索结果的概要信息。选择想要了解的记录, 如 HFE (Human Gene), 点击相应超链, 可获取其详细信息 (图 9-14), 包括三个方面的信息, 一是 HFE 基因总体信息, 如描述 (Description)、定位信息 (Location)、转录本信息 (Transcripts)。二是概要信息 (Summary), 如 CCDS、UniprotKB 蛋白质序列、RefSeq、Ensembl version、GRCh37 assembly、Gene type、Prediction Method、Alternative genes。三是采用图形显示该基因的综合注释。

The screenshot shows the Ensembl genome browser interface. At the top, there's a navigation bar with links like BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog, and Mirrors. A search bar on the right contains the text "Search all species". Below the navigation bar, the "Species" dropdown is set to "Human". The main search results are displayed in a table-like format. On the left, under "Restrict category to:", there's a list of categories and their counts: Gene (177), Transcript (254), Variation (2485), Phenotype (2), Somatic Mutation (85), Gene Tree (86), Genomic Alignment (493), Probe Feature (253), and Protein Family (220). Below this, under "Restrict species to:", there's a list of species and their counts: Human (741), Mouse (367), Zebrafish (50), Alpaca (12), and Amazon molly (11). The main content area shows the search results for "HFE". It indicates "4055 results match HFE". Below this, there are three main entries: "Hfe (Mouse Gene)", "HFE (Human Gene)", and "Hfe-004 (Mouse Transcript)". Each entry provides a brief description, a link to the gene page, and a link to the transcript page. For example, the "HFE (Human Gene)" entry mentions "Hemochromatosis [Source:HGNC Symbol;Acc:HGNC:4886] HFE (Vega gene) is associated with Gene ENSG0000010704".

图 9-13 Ensembl 的检索结果



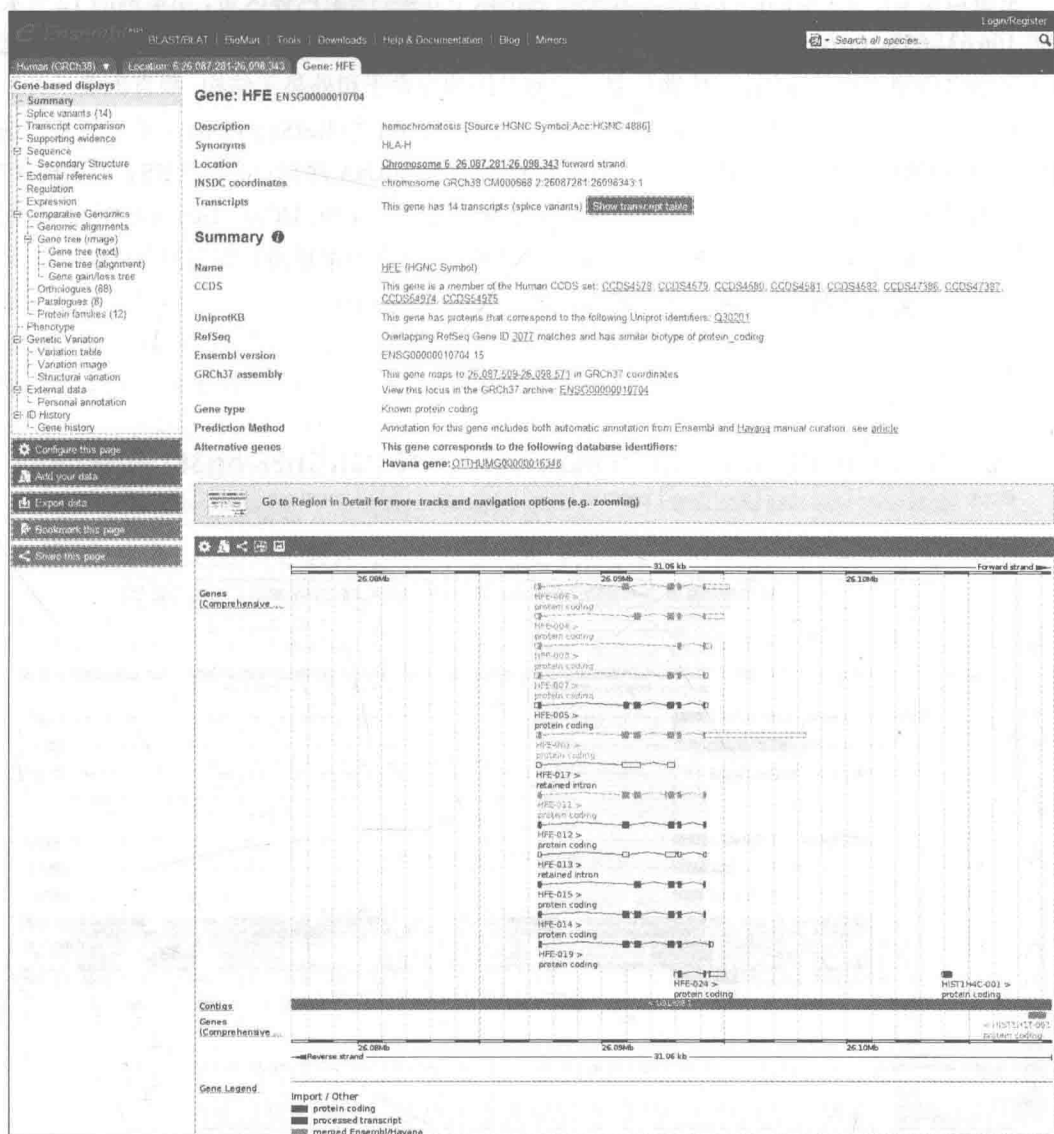


图 9-14 *HFE* 基因的详细信息

三、UCSC Genome Browser

UCSC Genome Browser (<http://genome.ucsc.edu/>) 是美国加利福尼亚大学 Santa Cruz 分校 (the University of California Santa Cruz, UCSC) 基因组信息学组建立的人类基因组图谱三大门户网站之一。该浏览器采用 NCBI 拼接整合的人类基因组序列作为平台, 提供了众多基因组中的定位数据, 包括染色体区带、连续子和间隙、mRNA 和表达序列标签 (EST)、预测基因、单核苷酸多态 (SNPs)、STS 的遗传和放射杂交图谱、重复序列、鼠同源序列、斑马鱼同源序列等。大多数数据与其他数据库链接, 如 RefSeq、LocusLink、PubMed、GeneLynx、GenCards、AceView 等。除了 NCBI Map Viewer 和 ensembl 以外, UCSC Genome Browser 也是获取人类和其他脊椎动物基因组信息的主要网络资源。

UCSC Genome Browser 支持关键词、基因名和其他文本查询,也可以用类似 BLAST 的比对工具 BLAT 进行数据检索。图 9-15 是查询细胞骨架相关蛋白 4 (cytoskeleton-associated protein 4,

CKAP4) 的检出结果。通过顶部的, 可以调节 DNA 的大小, 也可以点击染色体, 改变显示窗口。

主窗口是 *CKAP4* 基因的注释。注释主要包括以下几类: ①染色体区带, 如本例的 12 号染色体 106 631 659-106 641 713 区带; ② UCSC Genes: UCSC 基于 RefSeq、UniProt、GenBank 和比较基因组学预测出来的基因, 蓝色路径显示已知基因的内含子和外显子结构, 垂直框表示外显子而水平线则为内含子, 转录的方向由内含子上的箭头示意; ③ RefSeq Genes; ④来自基因编码第 19 版的比较基因组注释集; ⑤来自 GenBank 的人类 mRNA 序列; ⑥人类 EST 序列拼接; ⑦不同物种的多重比对; ⑧ DNA 重复片段。除了提供这些注释外, UCSC Genome Browser 允许用户为任何数据添加自己的注释。实际上, 在 UCSC 注释中一半是由外部用户贡献的。

窗口底部是一系列控制选项, 可以动态调整各类注释的显示。

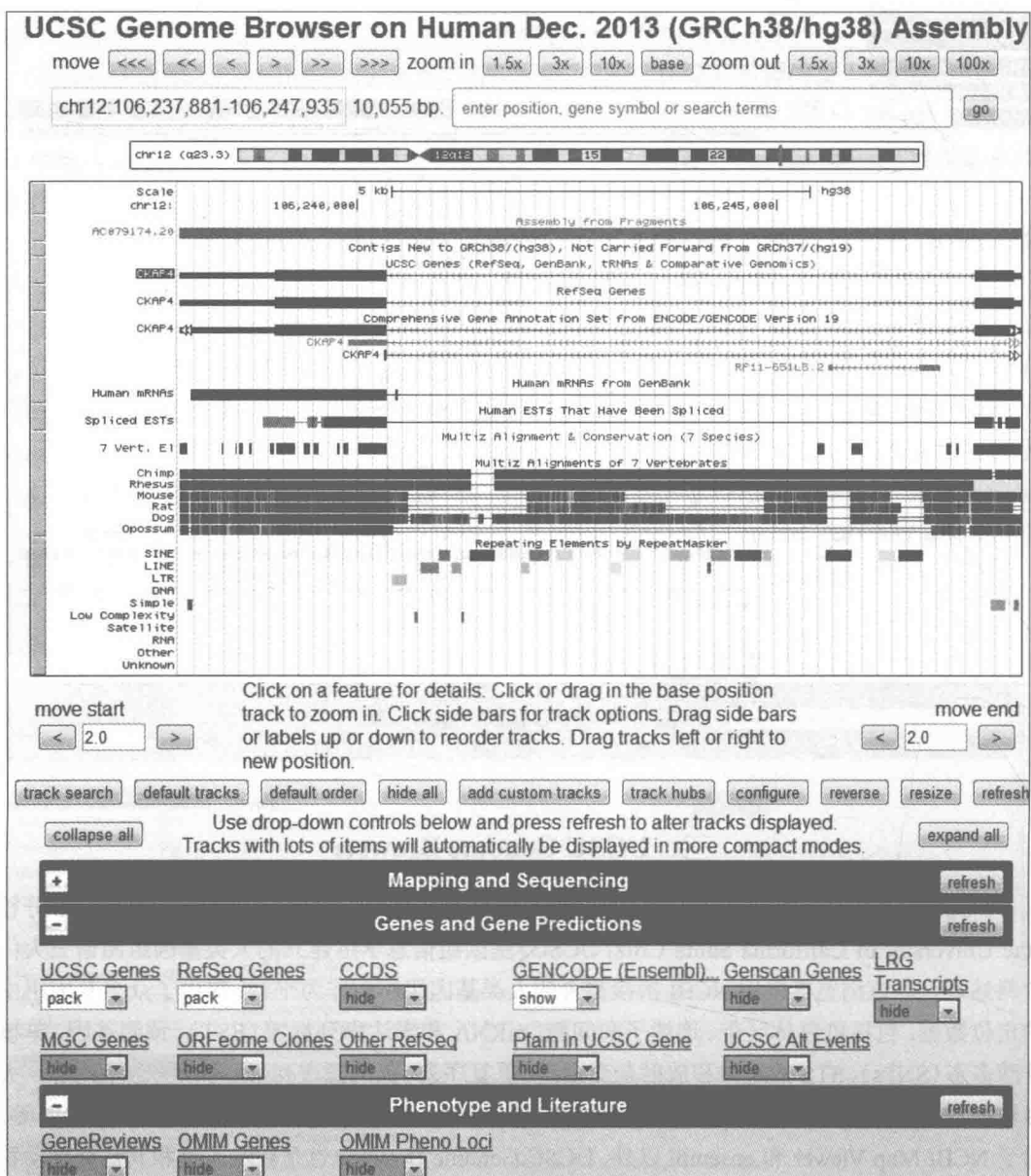


图 9-15 UCSC Genome Browser 检出结果 (CKAP4)



第五节 疾病基因数据库

一、OMIM

(一) OMIM 概述

在线人类孟德尔遗传数据库(Online Mendelian Inheritance in Man, OMIM, <http://www.omim.org/>), 是在美国 John Hopkins 大学医学院 Victor A. McKusick 教授编撰的《人类孟德尔遗传》一书的基础上发展起来的在线数据库, 内容包括所有已知的遗传病、遗传决定的性状及其基因, 除了简略描述各种疾病的临床特征、诊断、鉴别诊断、治疗与预防外等文本资料, 还提供已知有关疾病相关基因的连锁关系、染色体定位、组成结构和功能、动物模型, 并附有其相关的图片、研究历史和参考文献。OMIM 具有以下主要特点:

1. 及时性 它不仅是人类基因和遗传疾病的数据库, 更是不断更新的知识库, 内容几乎每天更新, 通过它可获知最新人类基因定位的信息。

2. 准确性 OMIM 的撰稿人具有良好的医学遗传学知识背景, 对于某一新发现的遗传性疾病要经过严格的审查、讨论, 方可收录, 保证了其准确性和权威性。

3. 全面性 OMIM 从 200 多种杂志发表的论文和专著中积累资料, 并增加新内容, 删除旧条目编辑而成。不仅可获取单基因病方面的资料, 而且可获取染色体病、多基因病、线粒体病方面的资料; 不仅列出了导致特定疾病的基因, 不少条目还列出了致病基因的等位变异体(allelic variants)。通过记录间的链接还可以查询 Medline、GenBank 等数据库。

4. 实用性 OMIM 中的文献经过专家的精选, 文献量远少于 PubMed, 可使科研人员从浩瀚的文献中解脱出来, 快速掌握某种疾病的遗传知识。如经 PubMed 检索, 1998 年至 2008 年有关 Marfan syndrome(马方综合征)的文献有 1684 篇, 而查 OMIM, 检出 80 条记录, 其中 #154 700(MARFAN SYNDROME), 列出了 132 篇重要的参考文献, 通过阅读这些文献, 对马方综合征从基础到临床就有了全面的了解。

(二) OMIM 数据的存贮

OMIM 每一条记录都有一个唯一的编号, 即 OMIM 存取号, 对应一个基因或者某种疾病。OMIM 存取号采用 6 位数字系统。如果存在等位基因, 则可以在 OMIM 存取号后添加四位小数数字用以区分不同的变异体, 例如血友病因子 IV 的所有变异体(或突变)可以分别命名为 306900.0001 到 306900.0101。此外, OMIM 存取号前通常带有一种前缀符号, 其含义及条目数量见表 9-2。

表 9-2 条目前缀及其含义和数量

前缀	含义	常染色体	X-连锁	Y-连锁	线粒体	合计
*	基因型描述	13 898	679	48	35	14 660
+	基因型和表型已知的基因	98	2	0	2	102
#	表型描述, 分子基础已知	3855	287	4	28	4174
%	表型或位点描述, 但是分子基础未知	1555	133	5	0	1693
	其他, 主要是尚未得以证实的孟德尔表型	1735	114	2	0	1851
	合计	21 141	1215	59	65	22 480

注: OMIM 条目数量来源于 <http://www.omim.org/statistics/entry> (2014 年 9 月 19 日更新)

(三) OMIM 检索

1. “OMIM”基本检索 检索方法与 GenBank 的 Entrez 检索相同(详见本章第二节), 可以

用记录内检索字段中出现的任何词(包括基因符号、疾病名称、症状、染色体定位、著者等)。词与词间可以用字段名称加以限定,也可用逻辑运算符相连进行复合检索。例如查找 10p21-22 区与哪些遗传疾病相关,可以输入染色体部位“10p21-22”,查到这一位置与猫眼综合征(cat eye syndrome)、精神分裂症(schizophrenia)、家族性脑膜瘤(familial meningioma)、前列腺癌(prostate cancer)等多种疾病有关。又如查 APC 基因相关的文献,在检索框内输入“APC”,可得 212 条记录。这些记录涉及结肠癌、胃癌、胰腺癌、食管癌等消化道有关的恶性肿瘤以及其他疾病的信息,还包括与 APC 相互作用的因子,如下调因子(APC, downregulated by, 1; APCDDI),并附有 PubMed 的链接。

OMIM 最常用的是对遗传性疾病的检索。如果要查先天性视网膜疾病的相关内容,可以输入“retina”,可得到 900 条记录。点击其中的一条如 +162 080,该记录是与视网膜有关的基因(neural retina leucine zipper),内容涉及该基因的描述(Description)、克隆(Cloning)、基因功能(Gene Function)、染色体定位(Mapping)、分子遗传学(Molecular Genetics)、动物模型(Animal Model)、等位基因变异(Allelic Variants)、参考文献(References)、编撰者(Contributors)、创建日期(Creation Date)、编辑历史(Edit History)等信息。

2. “Advanced Search(高级检索)” 提供了三个检索入口,第一个是 OMIM 的高级检索(图 9-16),即根据 OMIM 记录的字段特点,给出限制性检索字段的复选,以及染色体位置选择和数据时间段的限定。

图 9-16 OMIM 的高级检索

第二个是“clinical Synopsis”,临床症状高级检索,提供 inheritance(遗传)、growth(生长)、head and neck(头颈)、cardiovascular(心血管)、respiratory(呼吸)等 20 多个限定。

第三个是“Gene Map”,即基因图谱高级检索,它针对 OMIM 所描述的表达基因及疾病基因的细胞遗传学定位图(染色体定位)的检索。可以输入基因符号、染色体定位来检索(性染色体



X、Y 须大写), 如“CYP1”、“5”、“1pter”、“Xq”或“alzheimer”, 可限定在某一染色体、常染色体、表型。

(四) OMIM 在医学遗传学中的应用价值

1. 利用 OMIM statistics 了解最新的有关遗传病、性状及基因信息。

2. 利用 OMIM 数据库获得遗传病诊断、咨询及治疗的资料 由于发现和鉴定的遗传性疾病数目越来越多, 而遗传病大都较为罕见, 教学科研工作者、临床医师甚至医学遗传学专家所熟悉的病例毕竟有限, 而利用 OMIM 数据库就可以做到这点。在 OMIM 数据库击中某一条目后, 根据其中包含的 Clinical Features (临床特征)、Inheritance (遗传方式)、Diagnosis (诊断)、Clinical Synopsis (临床概要) 等栏目可快速地对病作出诊断, 查看 Clinical Management (临床治疗) 还可以获得先进的治疗方案。

3. 利用 OMIM 数据库获取遗传病最新、最详细的研究资料 利用 OMIM 数据库中的记录, 可以对某一遗传病从基础到临床作一个全面的了解, 如果对某一遗传病的某一具体方面如 Population Genetics (群体遗传学)、Cytogenetics (细胞遗传学)、Molecular Genetics (分子遗传学)、Mapping (基因定位)、Animal Model (动物模型) 等作深入的了解, 可进一步点击、阅读相关资料。

4. 利用 OMIM 的 Gene map 和 Morbid Map 获取疾病基因定位的详细资料 如输入疾病的名称, 就可查出该病的基因符号、基因位置信息; 输入基因符号可获知与该基因相关的疾病信息; 输入染色体区带号可获得该区带定位的基因及相关疾病信息。

二、GeneCards

GeneCards (<http://www.genecards.org/>) 是一个收集人类基因及其产物和相关疾病等综合信息的知识平台, 由以色列的 Weizmann 研究所基因组研究中心和生物信息学中心共同开发。内容包括基因命名及描述 (Aliases & Descriptions)、基因组定位 (Genomic Location)、蛋白质特征 (Protein Features)、蛋白质功能域及家族 (Protein Domains & Families)、基因功能 (Gene Function)、蛋白质通路和相互作用 (Pathways & Interactions)、药物及成分相关信息 (Associations with Drugs & Compounds)、转录本和剪接异构体 (Transcripts and Splice isoforms)、基因表达相关信息 (Expression-related information)、直系同源基因 (Orthologs)、旁系同源基因与假基因 (Paralogs & Pseudogenes)、单核苷酸多态性 (SNPs)、疾病与突变 (Disorders & Mutations)、医学新闻 (Medical News)、出版物 (Publications) 等。每个基因的信息来源于大约 50 个公共的和私有的数据库, 如 Ensembl、dbSNP、FlyBase、Genatlas、GenBank、GO、GeneAnnot、GeneLynx、GeneNote、GeneTests、HUGE、HUGO Gene Nomenclature Committee、OMIM、PubMed、Swiss-Prot、HGMD 等。到 2014 年 9 月, GeneCards 收录了 169 579 个基因信息, 其中 38 656 个已经被 HUGO 基因命名委员会审核通过, 疾病相关基因 6548 个。GeneCards 在网上免费提供简单检索 (Simple Search)、截词检索 (Wild Card Search)、多词逻辑组配检索 (Multiword Search) 和高级检索 (Advanced Search) 4 种检索方式。

三、GEO

1995 年 Schena 等在 *Science* 上发表了利用 cDNA 芯片研究拟南芥基因表达以及 1996 年 Affymetrix 的 Lockhart 等在 *Nature Biotechnology* 上发表了高通量 oligo 表达谱芯片, 开创了基因芯片研究的先河。此后, 基因芯片广泛用于表达谱分析、基因分型分析、比较基因组学分析、甲基化分析等多个领域。由此产生的海量数据需要数据库来储存和维护以方便世界各地的研究者查询和使用, 其中最重要的基因表达数据库是由 NCBI 开发维护的 GEO。

GEO (Gene Expression Omnibus, <http://www.ncbi.nlm.nih.gov/geo/>) 以芯片数据为主, 收录了基因表达谱数据、SNP 芯片数据、比较基因组学数据、microRNA 芯片数据等, 此外, 还收录

了一些非芯片类型的数据,如 SAGE(基因表达系列分析)数据、SARST(核糖体序列标签连续分析)数据、MS(质谱)数据、蛋白质组数据和新一代高通量测序数据(MPSS,大规模平行测序技术)等。GEO 对其收录的芯片技术平台的分类基于两个标准,一是芯片制备的技术,如点样(spotted)、原位合成(*in situ*)、磁珠(bead)等;二是探针的类型,如寡核苷酸(oligonucleotide)、cDNA 等。

GEO 的数据存储分成四个层次:① GPL-Platform:描述实验平台的信息,对芯片平台而言,描述芯片类型及芯片上的探针(probe)的数据,据此可以得知特定芯片包含多少个探针,每个探针对应着哪一个基因等信息;② GSM-Sample:描述单个芯片实验的杂交数据,据此可以得知特定样本与芯片杂交之后测得的 mRNA 表达量或 SNP 等位基因型等信息;③ GSE-Series:包含一系列的 GSM,样本间存在相关性,据此可以得知多个相关样本的所有基因的表达量等信息;④ GDS-Dataset:由 GEO 专家根据 GSM 编纂而成的条目,与 GSE 类似,描述了多个相关样本所有基因的表达量等信息。截至 2014 年 9 月,共包含 13 383 个 GPLs,1 236 744 个 GSMs 和 50 775 个 GSEs,以及 3847 个 GDS。

GEO 数据可以使用 Entrez GEO DataSets(数据集)和 Entrez GEO Profiles(表达谱)进行查询。Entrez GEO Profiles 查询预处理的基因表达/分子丰度图谱,即样品和系列记录,浏览器网址为 <http://www.ncbi.nlm.nih.gov/geo/query/browse.cgi>,而 Entrez GEO DataSets 查询所有的实验注解(<http://www.ncbi.nlm.nih.gov/geo/>),正如其他的 NCBI Entrez 数据库,可以使用布尔短语,并限定字段,进行有效的查询。

对于感兴趣的实验,使用 Entrez GEO DataSets 进行属性限定,如基因名、GEO 登录号、关键词、变异性、组织、创建日期和平台等。例如使用检索词 Expression profiling by array [DataSet Type] AND metastasis AND human [Organism],寻找有关人类新陈代谢的所有的核苷酸微点阵实验数据组。检索结果显示数据组标题、简短实验说明、分类法、实验变量类型和原始平台的链接、相关系列记录和完整 GDS 记录。一旦确定相关数据集,可进一步研究感兴趣基因的表达图谱。

Entrez GEO 表达谱进行属性限定,如关键词、平台和样品类型、提交者、组织、发表日期和补充文档类型等。例如利用检索策略 Type 1 diabetes [GDS Text] AND apolipoprotein [Gene Description] NOT Homosapiens [Organism],检索到所有在非人类的物种中 I 型糖尿病相关数据集中载脂蛋白相关的基因资料。检索结果显示报告人的注解、简短实验信息、分类法和这个图谱的条形索引图,单击索引图像可显示图谱的详细内容。

四、HGMD

HGMD 是由位于英国卡尔地夫医学遗传研究所(Institute of Medical Genetics in Cardiff, United Kingdom)构建的人类基因突变数据库(human gene mutation database, <http://www.hgmd.cf.ac.uk/ac/index.php>)。HGMD 从大约 250 种期刊中收集突变信息,用计算机和手工结合的方法来扫描这些期刊以寻找相关报道,包括在编码区、调控区和剪接区域的点突变,还包括插入、缺失、复制及重复,最近加入了疾病相关的多态。突变信息主要包括染色体定位、突变类型列表和相关的表型列表,并将基因内所有突变定位到 HGMD 的参考序列上,此外还链接到 OMIM、GDB 等相关数据库。至 2014 年 2 月,专业版 HGMD 收录了 6800 个基因中的 156 932 个突变。

五、单核苷酸多态性数据库

单核苷酸多态性(single nucleotide polymorphisms, SNPs)主要是指在基因组水平上由单个核苷酸的变异引起的 DNA 序列的多态性,是人类可遗传变异中最常见的一种,占有已知多态性的 90% 以上。SNPs 在人类基因组中广泛存在,平均每 100~300 个碱基对中就有 1 个,估



计其总数可达 300 万甚至更多。现代医学研究认为疾病的发生与基因变异密切相关。作为第三代基因多态性标志物, SNPs 研究对于阐明疾病的发病机制, 筛选对外界致病因素易感的易感人群等都具有重要意义。自 1998 年 SNP 数据库出现以来, SNPs 的总数现在已超过 350 万, 有 54 个模式生物的完整基因组测序已经完成, 另有 70 余种微生物的基因组正在测试之中。目前, 较为常用的 SNP 数据库主要有美国国家生物技术信息中心 (NCBI) 的单核苷酸多态性数据库 (dbSNP)、日本人群的 SNP 数据库 (JSNP) 等。

1. dbSNP 由美国国家生物技术信息中心 (NCBI) 和美国国家人类基因组研究所 (the National Human Genome Research Institute, NHGRI) 共同创建了单核苷酸多态性数据库 (dbSNP, <http://www.ncbi.nlm.nih.gov/snp>)。自 1998 年 9 月创建以来, dbSNP 一直都是一个主要的公共遗传变异数据库。dbSNP 与其他大型公共遗传变异数据库如 TSC 和 HGVbase 等都有信息的交换与整合。

dbSNP 单核苷酸多态性的类型: ①单个核苷酸的置换; ②小片段插入 / 切除多态; ③序列不变区; ④微卫星重复; ⑤命名变异 (named variants); ⑥特征不明确的杂合子分析。dbSNP 还包括疾病引起的临床性突变和中性的多态性。除了由提交者和 NCBI 共同指定的记录标识符外, dbSNP 录入的信息包括序列多态性、特定的实验条件、对携带这一变异的人群的描述以及人群或个体的基因型频率等资料。dbSNP 提供 Entrez 集成检索、限定检索和高级检索。

2. JSNP (a repository of Japanese Single Nucleotide Polymorphism data, <http://snp.ims.u-tokyo.ac.jp/>) 是日本人群的单核苷酸多态性数据库, 始建于 2000 年, 由东京大学医学科学研究所的人类基因组中心 (Human Genome Center, HCG) 和日本科技局 (Japan Science and Technology Agency, JST) 共同创建。JSNP 中每一个 SNP 的详细资料包括基本信息, 突变序列、已筛查 (screened) 序列、基因信息, 实验方法和 PCR 条件, 其中基本信息为 JSNP 发布的资料; 突变序列显示突变碱基及其两侧 5' 和 3' 端的 60 个碱基, 同时显示序列特征图和用于确定 SNP 的有效序列的数量, SNP 在基因中的实际位置显示于记录的序列浏览中; 基因信息包括基因索引和与 SNP 位置有关的注释信息。

本章小结

本章首先简要介绍了生物信息数据库类型、生物信息数据收集与存贮、生物信息数据库查找, 然后着重介绍了 GenBank、EMBL-Bank、DDBJ 三大核酸序列数据库, PIR、Swiss-prot 等蛋白质序列数据库, PDB、MMDB、DSSP、HSSP、SCOP、CATH 等蛋白质结构数据库, DIP、STRING、KEGG、BOND 等蛋白质功能数据库, Entrez Genomes、Ensembl、UCSC Genome Browser 等基因组数据库, OMIM、GeneCards、GEO 等疾病基因数据库。

(胡德华)

思考题

1. 生物信息数据库有哪些类型? 请举例说明。
2. 通过哪些途径查找所需的生物信息数据库?
3. 三大核酸序列数据库各有什么异同?
4. 蛋白质序列数据库有哪些? 各有什么特点?
5. PDB 数据库与 MMDB 数据库在蛋白质结构注释上有何异同?
6. 在 SCOP 数据库中, P53 抑癌基因 (p53 tumor suppressor) 属于哪个家族, 超家族和折叠类型?

7. 蛋白质功能数据库有哪些?
8. 利用 Entrez Genomes 的 Map Viewer, 查找 *RBP4* 基因定位于哪条染色体? 哪个区带?
9. 以 *ARMS2* 基因为例, 试比较 Entrez Genomes 的 Map Viewer、Ensembl 和 UCSC Genome Browser 三个基因组图谱浏览器在基因注释上的差异。
10. 利用 OMIM 等基因疾病数据库查找 *P53* 基因与哪些疾病存在关系, 并探讨它在这些疾病中的分子机制。

第十章 网络信息资源检索

进入 21 世纪,通过互联网获取各种信息已成为人们日常生活的一部分。对科技工作者来说,互联网更是重要的科研工具,通过它不仅可以了解科技发展的最新动态,获取最新信息资料,而且可以进行广泛的国际科技协作和学术交流。互联网上的信息资源极其丰富,利用好的搜索引擎可以帮助用户快速高效地获取信息,而开放存取资源则可以提供大量免费的学术文献。

第一节 网络信息资源概述

一、网络信息资源的定义

网络信息资源(information resource on network),是指通过计算机网络可以利用的各种信息资源的总和。具体来说是指所有以电子数据形式把文字、图像、声音、动画等多种形式的信息存储在光、磁等非纸介质的载体中,并通过网络通信、计算机或终端等方式再现出来的资源。

随着互联网发展进程的加快,信息资源网络化成为趋势,与传统的信息资源相比,网络信息资源在数量、结构、分布和传播的范围、载体形态、内涵和传递手段等方面都显示出了新的特点。目前网络信息资源以互联网信息资源为主,同时也包括其他没有连入互联网(局域网)的信息资源,其中包括生物学、医学、药学等生物医学信息资源。

网络学术资源是指与学术研究有关的互联网中的资源,主要包括政府信息、科研信息、教育信息和文化信息。政府信息是指国际组织、各国政府及其相关部门发布的信息;科研信息是专业学术机构,包括学会、协会、研究所等设立的网站及其相关信息;教育信息是大学设立的网站及相关信息;文化信息则是指各类信息媒体的网站、世界各地图书馆的数字化馆藏等。

二、网络信息资源的特点

随着网络技术的发展和网络浏览工具的开发,网络资源发展越来越快,网络信息资源数量大、表现形式多样化等特点愈发明显。

(一) 存储数字化

信息资源由传统纸张上的文字变为磁性介质上的电磁信号或者光介质上的光信息,使信息的存储、传递和查询更加方便,而且所存储的信息密度更高,容量更大,甚至可以无损耗地被重复使用。以数字化形式存在的信息,既可以在计算机内高速处理,又可以通过信息网络进行远距离传送。

(二) 表现形式多样化

传统信息资源主要是以文字或数字形式表现出来的信息,而网络信息资源则可以是文本、图像、音频、视频、软件、数据库等多种形式存在的,涉及领域从经济、科研、教育、艺术,到具体的行业或个体,包含的文献类型从电子报刊、电子工具书、商业信息、新闻报道、书目数据库、文献信息索引到统计数据、图表、电子地图等。

(三) 使用方便,共享性强

不同国家区域的人都可以利用网络传播信息,超级链接使查找信息灵活方便,组织形式多

样,信息获取快捷。网络信息资源还具有很强的交互性和共享性,人们可以就一个话题或研究课题展开交流和讨论,也可以让所有的网民共享自己的研究成果。

(四) 数量巨大,增长迅速

CNNIC(China Internet Network Information Center)一年发布两次的《中国互联网络发展状况统计报告》显示,截至2014年6月底,中国网民数量达到6.32亿,域名注册总量为1915万个,网站数量273万个,国际出口带宽总量为3 776 909Mbps。可见,互联网的极速发展推动了网络信息资源呈指数增长,数据庞大。

(五) 传播方式的动态性和时效性

网络环境下信息的传递和反馈快速灵敏,具有动态性和实时性等特点。信息在网络中的流动非常迅速,电子流取代了纸张和邮政的物流,加上无线电和卫星通信技术的充分运用,上传到网上的任何信息资源,都只需要短短的数秒钟就能传递到世界各地的每一个角落。

(六) 信息源复杂,管理困难

网络的共享性与开放性使得人人都可以在互联网上索取和存放信息,目前由于缺乏健全的质量控制和管理机制,这些信息无法经过严格编辑和整理,信息质量良莠不齐,大量不良和无用的信息充斥在网络上,形成了一个纷繁复杂的信息世界,给用户选择和利用网络信息带来了障碍。

第二节 搜索引擎

一、概 述

(一) 搜索引擎的定义

搜索引擎(search engine)是指根据一定的策略、运用特定的计算机程序搜集互联网上的信息,在对信息进行组织和处理后,将处理后的信息显示给用户,为用户提供检索服务的系统。它具有信息检索服务的开放性、超文本的多链接性和操作简单的特点。从使用者的角度看,搜索引擎提供一个包含搜索框的页面,在搜索框输入词语,通过浏览器提交给搜索引擎后,搜索引擎就会返回相关的信息列表。

(二) 搜索引擎的类型

搜索引擎按其工作方式主要分为三种,分别是全文搜索引擎、目录索引类搜索引擎和元搜索引擎。

1. 全文搜索引擎(full text search engine) 是名副其实的搜索引擎,国外具代表性的有Google,国内著名的有百度,它们都是从互联网上提取各个网站的信息(以网页文字为主)而建立数据库,从该数据库中检索与用户查询条件匹配的相关记录,然后按一定的排列顺序将结果返回给用户。全文搜索引擎的优点是数据库大、内容新、查询全面而充分,查全率高,能提供给用户全面而广泛的信息。缺点是查准率低,缺乏清晰的层次结构。

2. 目录索引类搜索引擎(search index/directory) 提供了一份人工按类别编排的网站目录,各类下边排列着属于这一类别网站的站名和网址链接,再记录一些对该网站进行概述性介绍的摘要信息(摘要可能是网站提交的,也可能是引擎站点的编辑为站点所做的评价)。用户搜索时就可以按相应类别的目录逐级浏览。这类引擎往往还伴有网站查询功能,也称之为网站检索,即提供一个文字输入框和一个用户可以在文字框中输入要查找的字、词或短语,再点击检索,便会在目录中查找相关的站名、网址和内容提要。目录索引中最具代表性的是Yahoo和新浪分类目录搜索。

目前,全文搜索引擎与目录索引类搜索引擎有相互融合渗透的趋势。



3. 元搜索引擎(META search engine) 也称集成搜索引擎,它没有自己的数据库,而是将用户的查询请求同时向多个搜索引擎递交,将返回的结果进行重复排除、重新排序等处理后,作为自己的结果返回给用户。这类搜索引擎的优点是返回结果的信息量更大、更全,缺点是不能够充分使用所使用搜索引擎的功能,用户需要做更多的筛选。著名的元搜索引擎有 InfoSpace、Dogpile、Vivisimo 等。

除上述三大类引擎外,还有几种非主流形式:集合式搜索引擎(collection of search engine)、门户搜索引擎(portal search engine)、免费链接列表(free for all links)等。

二、Google Scholar 和百度学术搜索

学术搜索引擎以网络学术资源为索引对象,一般涵盖互联网上的免费学术资源和以深层网页形式存在的学术资源,通过对这类资源的爬行、抓取、索引,以统一的入口向用户提供服务。学术搜索引擎把用户从面对海量、异构的学术资源不知如何下手的困境中解脱出来,它为用户屏蔽了不同数据源在平台、界面、检索指令等方面的差异,使用户可以通过一个简单的界面访问多种异构分布的资源。

(一) Google Scholar

Google 搜索引擎由美国斯坦福大学的 Larry Page 和 Sergey Brin 两位博士创建于 1998 年 9 月,是目前互联网上最大的搜索引擎。Google Scholar(<http://scholar.google.cn/>)是 Google 公司于 2004 年 11 月推出的一个免费学术搜索引擎。Google Scholar 的资料来源主要有下面几个方面:一是网络免费的学术资源,包括论文预印本、会议论文、调研报告等;二是开放存取(详见本章第三节)的期刊网站;三是付费电子资源供应商,例如我国的维普数据库就与他们合作,提供中文期刊文献的题录信息;四是图书馆链接,Google 与图书馆合作,在搜索结果中将包含一个到图书馆相关资源的链接。

Google Scholar 的检索功能灵活、强大,尤其是支持多种字段检索、特定文件类型检索等功能,并可以按用户的习惯设置检索界面。

(二) 百度学术搜索

百度搜索引擎是全球最大的中文搜索引擎之一,2000 年 1 月由李彦宏、徐勇两人创立于北京中关村,致力于向人们提供“简单,可依赖”的信息获取方式。“百度”二字源于中国宋朝词人辛弃疾的《青玉案》诗句:“众里寻他千百度”。百度搜索引擎使用超链接技术即通过分析链接网站的多少来评价被链接的网站质量,让用户在百度搜索时,越受用户欢迎的内容排名越靠前。

2014 年 6 月,百度学术搜索(<http://xueshu.baidu.com>)上线。百度学术搜索是百度旗下的提供海量中英文文献检索的学术资源搜索平台,涵盖了各类学术期刊、会议论文等资源。百度学术搜索可检索到收费和免费的学术论文。在百度搜索页面下,会针对用户搜索的学术内容,呈现出百度学术搜索提供的合适结果。用户可以选择查看学术论文的详细信息,也可以选择跳转至百度学术搜索页面查看更多相关论文。还可以选择将搜索结果按照“相关性”、“被引频次”、“发表时间”三个维度分别排序,以满足不同的需求。

三、医学搜索引擎

(一) Medical Matrix

1994 年由美国医学信息学会主办,是目前最重要的医学专业搜索引擎之一。它是一种经评估的主题目录式网上医学信息搜索工具,即只有经过美国医学信息学会的资深专家评估后被认为有价值的网站信息才被网站收录,所以访问者很容易利用 Medical Matrix 网站找到符合自己要求又具有权威性的信息。Medical Matrix 是一种由概念驱动的免费全文智能检索工具,主要服务于从事临床及卫生专业工作的用户。

Medical Matrix (<http://www.medmatrix.org/index.asp>) 提供了分类检索和关键词检索两种搜索方式。分类检索是它的主要特色,将各种医学信息分为专业(Specialties)、疾病(Diseases)、临床实践(Clinical Practice)、文献(Literature)、教育(Education)、卫生保健和职业(Healthcare and Professionals)、医学计算互联网和技术(Medical Computing, Internet and Technology)、市场(Marketplace)八大类。每一大类下再根据资源的组织形式及具体情况又进一步划分为检索(Searches)、新闻(News)、全文和多媒体(Full Text/Multi Media)、期刊(Journals)、教学参考书(Textbooks)、主要网址(Major Sites/Home Pages)、操作指南(Practice Guidelines/FAQs)、影像学和病理切片/Images, Path/Clinical)、医学继续教育(CME)、患者教育(Patient Education)、名录(Directories)、教育资源(Educational Materials)、论坛(Forums)等亚类。

Medical Matrix 收集的内容专业、全面,而且对每一内容均有简明扼要的评论,并且对链接的网址按一到五个星进行分级,便于使用者事先决定是否进入其网页进一步阅读,以节省时间。评价指标为:1星(1~10分)临床内容适合,较权威的临床医学信息,但缺少主旨和内容;2星(11~20分)临床内容较可靠、定期更新、站点设计合理且容易使用,但作为一个常用的临床医学信息资源站点功能有限;3星(21~30分)权威、准确、临床信息内容新、站点设计好、维护较好、功能全面、对一般用户而言检索容易、方便导航,提供的是有价值的临床医学信息资源;4星(31~40分)包含本学科所有类目和 Web 页面的知名站点;5星(41~50分)获得奖励的医学站点。关键词检索是在“Search for”检索框内输入关键词或短语,从而获得相应信息。关键词检索支持模糊查询及通配符(*)查询。输入短语时,可以选择匹配方式:精确匹配、所有的检索词(逻辑与)、任意的检索词(逻辑或);查询的选择范围是:整个站点、新闻、影像学和病理切片、X线影像、患者教育、CME、药物分类。在高级检索方式中,还可以限定查询内容的类型(如参考书、实践指南等),可以单选或多选,结果显示命中的相应链接或站点。

(二) Medscape

由美国 Medscape 公司 1994 年开发,1995 年 6 月投入使用,是互联网上最大的免费提供临床医学全文文献、药物数据库和医学继续教育资源(CME)的站点,主要为临床医生和其他医学工作者提供高质量的、及时的专业医学信息。

Medscape 主页(<http://www.medscape.com>)的上部有 News & Perspective、Drug & Diseases 和 CME & Education 三个标签,可以进行相应的浏览也可以进行检索。网页顶部的 MULTISPECIALTY 包含 30 多个专业,几乎所有生物医学工作者都可以找到相对应的专业主页,每一个专业都提供该专业的相关信息及深度报道。CME 是 Medscape 中最值得关注的内容,目前网上有超过 300 个继续教育课程可按专业查找,皆为近年来的医学论文。在阅读完每篇文章后可立即在网上进行测试。若能答对 70% 的试题就可获得美国医学会认可的继续教育学分(网上有追踪软件可自动记录浏览时间及测试分数,并可实时打印出学分证明)。

Medscape 提供免费注册功能,用户免费注册后,Medscape 根据用户登记的专业提供最新的来自路透社、专业期刊的信息,也可以免费订阅每周期的精选信息 MedPulse,还可以根据个人的需要定制个人独特的 Medscape 的界面。

(三) HON

HON(http://www.hon.ch/index_cn.html)是“健康在线基金会(Health On the Net Foundation, HON)”于 1996 年 3 月推出的一个检索型医学免费全文搜索引擎,该网站提供的医学信息可靠,为一些不懂医的人、医学信息使用者和医学专业人员提供了可靠的信息来源,成为最受欢迎的互联网上非营利性门户网站。

HON 分别提供英语、法语、德语、中文、西班牙语、葡萄牙语和荷兰语服务,指导用户获得有用且可信的医疗与健康信息。HON 的资源针对人群分为普通病人、医务工作者和网站管理员。HON 主页还提供 HONcode 网站、医学网站、HONselect、新闻、医学会议、医学图像、



HON 档案、HON 项目、WRAPIN (HON 领导的旨在提供网上医学信息质量监控的欧盟项目) 等栏目供用户浏览, 包含的信息量非常大。HON 推出免费全文医学搜索引擎 MedHunt 和利用 MedHunt 完成检索的 HONselect 服务项目, 将两个检索工具整合在同一页面中供用户使用。

HONselect (<http://www.hon.ch/HONselect>) 是整合后新的医疗与健康信息搜索引擎, 将 MeSH 词、权威科学论文、医疗信息、Web 地址和多媒体五种信息搜索集成为一站式搜索服务, 以便集中和加快信息搜索。

(四) Medical World Search

Medical World Search (<http://www.Mwsearch.com>) 是美国 The Polytechnic Research Institute 1997 年建立的。采用统一医学语言系统, 融合了 30 余种生物医学词表和分类法, 可以使用 500 000 多个医学术语, 免费全文检索, 免费检索 MEDLINE, 可以自动调整检索词, 使其符合其他检索工具如 Hotbot、infoseek、AltaVista 等的要求。

四、搜索引擎的使用技巧

(一) 检索基础使用技巧

1. 搜索关键词提炼 选择合适的关键词是成功检索的第一步。选择搜索关键词的原则是, 首先确定目标, 形成一个比较清晰的概念, 即要找的到底是什么? 是资料性的文档? 还是某种产品或服务? 然后再分析这些信息都有哪些共性, 以及区别于其他同类信息的特性, 最后从这些方向性的概念中提炼出此类信息最具代表性的关键词。如果这一步做好了, 往往就能迅速地定位要找的东西, 而且多数时候根本不需要用到其他更复杂的搜索技巧。关键词的选择虽然没有什么定式, 不过也有一些规律可循, 在提取关键词时力求做到规范, 避免口语化, 具体要求如: 使用特定词汇、使用多个含义相近的关键词等。

2. 细化搜索条件, 确定网站 搜索条件越具体, 搜索引擎返回的结果也会越精确。许多单位都会针对同一热门事件在本单位网站发大量文件或文章。这些文章或文件往往名称相同但具体内容和要求却不同。此时, 指定网站或频道的搜索功能能为我们提高搜索效率。

3. 按文件类型查找 如需要查找简历模板时可以构建如下检索式: “简历模板 filetype: doc”, 此时能搜到 word 格式的简历模板。此处的 filetype: doc 是限定检索文件的类型, 该字段必须在英文状态下录入, 而“简历模板”是检索的关键词。两个字之间必须用空格隔开。同理, 如果需要检索市场营销的教案也可以构造“市场营销教案 filetype: ppt”的检索式, 检索结果为 ppt 格式的市场营销教案。该检索式支持 doc、txt、ppt、xls、pdf、mp3、swf 等常见文件类型。

4. 巧用快照 有时会搜到访问不了的“死链网页”和过期文件, 而看其内容摘要很符合需要。此时可利用搜索引擎提供的网页快照功能查阅此类文献。

5. 使用不同的搜索引擎 不同的搜索引擎, 其信息覆盖范围有差异, 搜索信息时仅集中于某一家搜索引擎是不明智的, 因为再好的搜索引擎也有局限性, 而使用多个搜索引擎查询可以达到事半功倍的效果。

6. 留意搜索引擎返回的结果 由于竞价排名业务的推广, 搜索引擎返回的 Web 站点顺序会影响人们的访问。所以, 为了增加 Web 站点的点击率, 一些 Web 站点会付费给搜索引擎, 在相关 Web 站点列表中显示在靠前的位置。好的搜索引擎会鉴别 Web 站点的内容, 并据此安排它们的顺序, 但许多搜索引擎无法鉴别。

(二) 提高查询技巧

在搜索引擎中输入关键词, 然后点击“搜索”, 系统快速返回查询结果, 这是最简单的查询方法, 使用方便。但是简单查询的结果常常不准确, 可能包含着许多无用的信息, 此时需要运用命令符号大幅提高我们的搜索精度。

1. 使用双引号 (" ") 给要查询的关键词加上双引号(半角, 以下要加的其他符号同此), 可

以实现精确的查询,这种方法要求查询结果要精确匹配,不包括演变形式。例如在搜索引擎的文字框中输入“电传”,它就会返回网页中有“电传”这个关键字的网址,而不会返回诸如“电话传真”之类网页。

2. 使用加号(+) 在关键词的前面使用加号,也就等于告诉搜索引擎该单词必须出现在搜索结果中的网页上,例如,在搜索引擎中输入“电脑+电话+传真”就表示要查找的内容必须要同时包含“电脑”、“电话”、“传真”这三个关键词。

3. 使用减号(-) 在关键词的前面使用减号,也就意味着在查询结果中不能出现该关键词,例如,在搜索引擎中输入“电视台-中央电视台”,它就表示最后的查询结果中一定不包含“中央电视台”。

4. 使用通配符(*和?) 通配符包括星号(*)和问号(?),前者表示匹配的数量不受限制,后者匹配的字符数要受到限制,主要用在英文搜索引擎中。例如输入“computer*”,就可以找到“computer、computers、computerised、computerized”等单词,而输入“comp?ter,则只能找到“computer、compater、competer”等单词。

5. 使用布尔检索 所谓布尔检索,是指通过标准的布尔逻辑关系来表达关键词与关键词之间逻辑关系的一种查询方法,这种查询方法允许我们输入多个关键词,各个关键词之间的关系可以用逻辑关系词来表示。and,称为逻辑“与”,用 and 进行连接,表示它所连接的两个词必须同时出现在查询结果中,例如,输入“computer and book”,它要求查询结果中必须同时包含 computer 和 book。or,称为逻辑“或”,它表示所连接的两个关键词中任意一个出现在查询结果中就可以,例如,输入“computer or book”,就要求查询结果中可以只有 computer,或只有 book,或同时包含 computer 和 book。not,称为逻辑“非”,它表示所连接的两个关键词中应从第一个关键词概念中排除第二个关键词,例如输入“automobile not car”,就要求查询的结果中包含 automobile(汽车),但同时不能包含 car(小汽车)。near,它表示两个关键词之间的词距不能超过 n 个单词。在实际的使用过程中,可以将各种逻辑关系综合运用,灵活搭配,以便进行更加复杂的查询。

6. 使用括号 当两个关键词用另外一种操作符连在一起,而又想把它们列为一组时,就可以对这两个词加上圆括号。

7. 使用特殊搜索命令 大多数搜索引擎都支持针对网页标题的搜索,命令是“title:”,如在搜索引擎中输入“title:清华大学”,就可以查到网页标题中带有清华大学的网页。在进行标题搜索时,前面提到的逻辑符号和精确匹配原则同样适用。此外还可以针对网站进行搜索,命令是“site:”(Google)、“host:”(AltaVista)、“url:”(Infoseek)或“domain:”(HotBot)。在 Google 和 AltaVista 中,用户均可通过“link:”命令来查找某网站的外部导入链接(inbound links)。其他一些引擎也有同样的功能,只不过命令格式稍有区别。

8. 区分大小写 这是检索英文信息时要注意的一个问题,许多英文搜索引擎可以让用户选择是否要求区分关键词的大小写,这一功能对查询专有名词有很大的帮助,例如:Web 专指万维网或环球网,而 web 则表示蜘蛛网。

第三节 开放存取资源

一、概述

开放存取(open access, OA)是国际科技界、学术界、出版界、信息传播界为推动科研成果,利用网络自由传播而发起的运动。1991年互联网上出现第一个OA知识库arXiv,其初衷是解决当前的学术期刊出版危机,促进学术信息的交流与出版,提升科学研究的公共利用程度,保

障科学信息的长期保存。

2001 年 12 月,开放协会研究所在匈牙利的布达佩斯召集了一次有关开放访问的国际研讨会,并起草和发表了“布达佩斯开放存取倡议”(Budapest Open Access Initiative, BOAI)。按照 BOAI 中的定义,开放存取是指某文献在 Internet 公共领域里可以被免费获取,允许任何用户阅读、下载、拷贝、传递、打印、检索、超级链接该文献,并为之建立索引,用做软件的输入数据或其他任何合法用途。用户在使用该文献时不受财力、法律或技术的限制,而只需在存取时保持文献的完整性,对其复制和传递的唯一限制,或者说版权的唯一作用应是使作者有权控制其作品的完整性及作品被准确接受和引用。这是一种新的学术信息交流的方法,作者提交作品不期望得到直接的金钱回报,而是提供这些作品使公众可以在公共网络上利用。

开放存取有两种具体形式。一种是开放存取期刊,是依托网络技术,采用发表付费,阅读免费的形式,即出版物的费用由作者支付,而不是由图书馆订购后让读者获取,任何人可以在图书馆没有购买的情况下免费获取开放出版期刊上的论文全文;另一种形式是作者自建文档向外界免费开放,即作者将自己的研究成果以电子全文形式存放在一个中心服务器或 Web 网页上供同行免费利用的一种方式。

开放存取的信息是一种数字化的、在线的、大部分情况下在获取时不受版权与注册限制的文献,并具有时效性强、交互性强和学术价值高的特点。例如电子文本文献(eprints)不经过出版审查程序,作者的手稿可以在最短的时间内向网络公开,省去了一般出版物的出版周期,能够反映最新的研究成果与动态,这对科研人员及时了解专业动态尤为重要。许多开放存取期刊与传统出版刊物一样须经过严格的同行评审才能出版,保证了开放存取期刊的学术质量。目前著名的开放存取期刊来源,主要集中在生物和医学领域。

开放存取的实现模式主要包括:机构知识库(如 D-space、Euclid、D-Scholarship、KU ScholarWorks)、学科知识库(如 PubMed Central、arXiv.org 预印本资源库)和开放存取期刊,一些机构还专门致力于开放存取期刊的出版(如 BioMed Central)。

2014 年 5 月中国科学院和国家自然科学基金委分别发布了实行开放获取的政策声明,《国家自然科学基金委员会关于受资助项目科研论文实行开放获取的政策声明》中写到:国家自然科学基金全部或部分资助的科研项目投稿并在学术期刊上发表研究论文的作者应在论文发表时,将同行评议后录用的最终审定稿,存储到国家自然科学基金委员会的知识库,不晚于发表后 12 个月开放获取。《中国科学院关于公共资助科研项目发表的论文实行开放获取的政策声明》中指出:我院研究人员和研究生以我院所属机构名义承担的各类公共资助科研项目,自本政策发布之日起投稿并在学术期刊上所发表的研究论文,作者应在论文发表时把同行评议后录用的最终审定稿存储到所属机构的知识库,并于发表后 12 个月内开放获取。

二、国外主要开放存取资源

(一) HighWire Press

HighWire Press (<http://home.highwire.org/>) 是斯坦福大学著名的学术出版商,目前已成为世界三个最大的能够联机提供免费学术论文全文的出版商之一。截至 2014 年 8 月底,HighWire Press 提供免费检索的期刊 3546 种,共收录近 234 余万篇免费全文,包括物理、生物、医学和社会学领域的核心期刊论文。通过 HighWire 网站还可以检索 MEDLINE 的文摘题录。

HighWire Press 为出版商、图书馆员和研究者提供不同的服务页面。研究者可以从期刊字顺(By Title)、出版社(By Publisher)、期刊主题(By Topic)等途径浏览期刊列表。选中某刊后,点击刊名链接可以进入该刊页面进行浏览,如果在文章后标出可获取全文,则可以得到该文献的 PDF 格式全文。供免费浏览的期刊有三种:一是完全免费(free site),二是免费试用(free trial),三是部分免费(free issues)。

HighWire Press 检索功能较强(图 10-1),可以输入作者、关键词、年、卷、页等限制进行检索,关键词输入框右侧的三个选项: any 表示输入的两个或两个以上的词之间逻辑关系为“或”, all 表示输入的两个或两个以上的词之间逻辑关系为“与”, phrase 表示输入的两个或两个以上的词作为短语检索。检索的范围用户还可以有三个选择: HighWire-hosted content only 表示只在 HighWire 出版社的期刊中检索, Include PubMed 表示同时也检索 PubMed 的数据, My Favorites only 表示在用户自己定制的期刊中检索(定制服务可以免费注册后开通)。此外,用户还可将检索结果限定为综述文献,还可以选择出版时限。点击检索结果后的文摘、全文、期刊网站、引用图链接,可以分别获得文摘、全文(PDF 格式)、该期刊的网站主页和该文被引用的情况等。



图 10-1 HighWire 研究者检索界面

(二) DOAJ

DOAJ(Directory of Open Access Journals, <http://www.doaj.org/>)是由瑞典隆德大学图书馆主办、OSI 和 SPARC 协办整理的开放存取期刊目录,建于 2003 年 5 月。截至 2014 年 8 月底,网站共收录期刊 9966 种,收录开放存取文章 170 余万篇。DOAJ 收录的期刊均为学术性、研究性的同行评审期刊,或者有编辑做质量控制的期刊,具有免费、提供全文、高质量的特点,对学术研究有很高的参考价值。收录与生物医学相关的主题有:农业及食品科学、生物及生命科学、化学、地球及环境科学、健康科学等。

DOAJ 遵循创作共用约定,该目录及其收录期刊、论文可自由存取,任何人可以使用,不反对商业用途,目的也是为了改善学术期刊的可见性与可用性,增加学术文章的影响力。DOAJ 论文存贮格式以 HTML、PDF 为主,所有组成部分,无论是作者引文还是正文,用户都可以免费方便直接地进行浏览、下载和打印。

DOAJ 的检索可在所有字段、题名、期刊名、ISSN、作者名、关键词、文摘等字段中检索,在检出结果列表中点击文献题录下方的全文链接网址即可打开全文。

(三) BioMed Central

BioMed Central(BMC, <http://www.biomedcentral.com>)是一家英国商业性出版机构,是重要



的开放存取期刊出版商之一。目前出版 265 种经同行评审的开放获取期刊,内容涵盖生物学和医学的所有主要领域。

点选主页面(图 10-2)上的“Journals”用户可以通过 Journals A~Z 或 Browse by subject 按字顺或主题浏览期刊列表,再选择期刊阅读全文,亦可通过点选主页面上的“Articles”直接浏览期刊论文。主页面有检索框供快速检索,用户也可以点击“Advanced search”进入高级检索界面,进行多字段布尔逻辑检索。



图 10-2 BioMed Central 主页

(四) PLOS

公共科学图书馆(The Public Library of Science, PLOS, <http://www.plos.org/>)是一家由众多诺贝尔奖得主和慈善机构支持的非营利性学术组织,旨在推广全世界的科学和医学领域的最新研究成果。2000 年 10 月, PLOS 由生物医学科学家 Harold E. Varmus、Patrick O. Brown 和 Michael B. Eisen 创立,它的第一个行动就是发布一封公开信,鼓励科学出版商把研究文献通过在线免费公共文件发布(例如美国国家医学图书馆的 PubMed 中心)。这封公开信来自大约 180 个国家的 34 000 名科学家签署,加速了许多科学出版商更自由地获得已发表研究途径的脚步。2003 年, PLOS 作出一个决定,非营利地为科学家和医生提供高质量高水平的期刊,并会发布他们最重要的作品。在这个开放资源的模式下, PLOS 期刊直接在网上可以看到,并可免费使用,之后再发布或使用也没有任何限制,只要按创作共享注明出处授权条款的要求注明作者和来源即可。

目前 PLOS 出版 7 种生命科学与医学期刊,它们是: PLOS Biology、PLOS Medicine、PLOS Computational Biology、PLOS Genetics、PLOS Pathogens、PLOS ONE、PLOS Neglected Tropical Diseases。用户既可以浏览某一种期刊,并在此刊范围内检索文献,也可以在 7 种期刊中检索文献。

(五) Free Medical Journals

Free Medical Journals (<http://www.freemedicaljournals.com>)是由法国的 Bernd Sebastian Kamps 建立的免费医学期刊信息网站,是互联网上免费提供生物医学全文最多的期刊集合网站,截至

2014年8月底该网站共收录4308种生物医学期刊。该网站提供四种浏览途径,一是按学科专业浏览,将所有免费期刊分成89个专业,每一专业后在括号内用数字表明“该专业ISI有影响因子的期刊数/该专业免费期刊总数”;二是按FMJ自己的影响因子浏览期刊;三是按开放获取的时间选择期刊浏览,分为即时开放、1~6个月开放、7~12个月开放和更长时间开放四种;四是按刊名字顺浏览。该网站还提供FreeBooks4Doctors的链接以便用户免费阅读开放存取图书。

(六) PubMed Central

PubMed Central (PMC, <http://www.ncbi.nlm.nih.gov/pmc/>) 是2000年1月由美国国家医学图书馆(NLM)的国家生物技术信息中心(NCBI)建立的生命科学期刊全文数据库,它旨在保存生命科学期刊中的原始研究论文的全文,并在全球范围内免费提供使用。PMC采取自愿加入的原则,某期刊一旦加入,必须承诺期刊出版后一定时期内将其全文提交给PMC,由PMC提供免费全文检索和访问。PMC与PubMed的关系:两者都是NLM建立的数据库,其中PubMed是一个基于互联网的文献检索系统,它收录了几千种生命科学期刊的目次和文摘,该数据库提供了与PMC全文的链接以及数千种期刊网站的链接。而PMC是由NLM建立的免费生命科学电子期刊全文数据库,PMC的所有论文在PubMed中都有相应的记录。

PMC的界面与PubMed类似,除了可以浏览期刊列表外,还可以进行基本检索、高级检索和限定检索。

三、国内主要开放存取资源

(一) 中国科技期刊开放获取平台

中国科技期刊开放获取平台(China Open Access Journals, COA, <http://www.oaj.cas.cn>)由中国科学院主管,是一个开放获取、学术性、非营利的科技文献资源门户,于2010年10月上线运行。截至2014年8月底COA共收录832种期刊,其中医药卫生期刊有146种。

该平台(图10-3)提供期刊检索和论文检索两个途径,用户也可以通过期刊目录直接查找、浏览期刊。



图 10-3 中国科技期刊开放获取平台主页

（二）中国科技论文在线

中国科技论文(<http://www.paper.edu.cn/>)在线是经教育部批准,由教育部科技发展中心主办的科技论文网站。该网站打破传统出版物的概念,免去传统的评审、修改、编辑、印刷等程序,给科研人员提供一个方便、快捷的交流平台,提供及时发表成果和新观点的有效渠道,从而使新成果得到及时推广,科研创新思想得到及时交流。根据文责自负的原则,只要作者所投论文遵守国家相关法律,为学术范围内的讨论,有一定学术水平,且符合中国科技论文在线的基本投稿要求,可在一周内发表。中国科技论文在线可为在该网站发表论文的作者提供该论文发表时间的证明,并允许作者同时向其他专业学术刊物投稿,以使科研人员新颖的学术观点、创新思想和技术成果能够尽快对外发布,并保护原创作者的知识产权。平台提供浏览和快速检索、高级检索功能。

（三）中国预印本服务系统

预印本(preprint)是指科研工作者的研究成果还未在正式出版物上发表,而出于和同行交流目的自愿先在学术会议上或通过互联网发布的科研论文、科技报告等文献。与刊物发表的文章以及网页发布文章比,预印本具有交流速度快、利于学术争鸣、可靠性高的特点。

中国预印本服务系统(<http://prep.istic.ac.cn>)是由中国科学技术信息研究所与国家科技图书文献中心联合建设的,以提供预印本文献资源服务为主要目的的实时学术交流系统。该系统主要收藏的是国内科技工作者自由提交的预印本文章,可以实现文献检索、分类浏览等功能。系统收录的预印本内容主要是国内科技工作者自由提交的科技文章,一般只限于学术性文章。系统的收录范围按学科分为五大类:自然科学、农业科学、医药科学、工程与技术科学、人文与社会科学,系统不对文章进行学术审核。

（四）开放阅读期刊联盟

开放阅读期刊联盟(<http://www.oajs.org/>)是国内几家重点大学学报发起的开放阅读项目。包括西安交通大学学报自然版、西安电子科技大学学报自然版、清华大学学报自然版、东南大学学报自然科学版等数十家学报期刊。这些大学学报网站上提供全文免费供读者阅读,或者应读者要求,在3个工作日之内免费提供各自期刊发表过的论文全文(一般为PDF格式)。读者可以登录各会员期刊的网站,免费阅读或索取论文全文。目前收录的医学类期刊有:西安交通大学学报医学版、第四军医大学学报、南方医科大学学报、成都医学院学报、实用医学杂志。

（五）中国学术会议在线

中国学术会议在线(<http://www.meeting.edu.cn/>)是经教育部批准,由教育部科技发展中心主办的面向广大科技人员的科学研究与学术交流信息服务平台。中国学术会议在线利用现代信息技术手段,分阶段实施学术会议网上预报及在线服务、学术会议交互式直播/多路广播和会议资料点播三大功能。为用户提供学术会议信息预告、会议分类搜索、会议在线报名、会议论文征集、会议资料发布、会议视频点播、会议同步直播等服务。

四、慕课资源

（一）网络开放课程

开放课程是指在网络上传播的国内外优秀大学的视频课程,又名公开课。MOOC(massive open online course)是2012年以来迅速崛起的一种新型教育模式,也称为大规模开放在线课程或者大规模网络公开课,也有人称之为慕课。“大规模”意味着对学习人数没有限制,“开放”则意指公众可免费注册学习。与视频公开课不同之处在于,MOOC具有完整的教学过程,有固定的开课时间,课程内容非结构化,视频多在10~20分钟有作业和考试环节,更有教学双向互动和学习者之间的交流。

(二) 国外主要慕课资源

1. Coursera (<https://www.coursera.org>) 是免费大型公开在线课程项目,由美国斯坦福大学两名计算机科学教授安德鲁·恩格(Andrew Ng)和达芙妮·科勒(Daphne Koller)于2012年4月创办。Coursera的首批合作院校包括斯坦福大学、密歇根大学、普林斯顿大学、宾夕法尼亚大学等美国名校。截至2014年8月已有来自于世界110所大学的733门课程可供学习。课程涵盖生命科学、食物和营养学、医学等学科。

2. edX (<https://www.edx.org>) 是2012年由美国麻省理工学院和哈佛大学联手创办的非盈利网络教育平台,向全球免费提供知名高校的优质课程。截止到2014年8月已有麻省理工学院、哈佛大学、加利福尼亚大学伯克利分校、德克萨斯大学系统、北京大学、清华大学等42所高校或机构的200余门课程可供学习,学科内容涵盖生物学、化学、计算机科学、工程、食物与营养、人文学科、医学等。

3. Udacity (<https://www.udacity.com>) 提供美国斯坦福大学公开课程。Udacity最早起源于斯坦福大学的两位教师 Sebastian Thrun 和 Peter Norvig 提供的在线免费网络课程《人工智能导论》。截至2014年8月该项目提供39门课程供学习,包括经济、计算机、设计、数学和生物医学等学科,分为初级、中级和高级三个层次。

(三) 国内主要慕课资源

1. 国家精品课程资源网 (<http://www.jingpinke.com>) 是由国家教育部主导推动的国家级精品课程集中展示平台,汇集了万余门各学科、专业精品课程,课程内容涵盖全国本科和高职院校各级精品课程、国家精品课程名师主讲的精品课程建设培训课程、欧美亚高等学府的4000多门开放课程、国内外知名企业优质课程。课程主要涵盖文学艺术、哲学历史、经管法学、基础科学、工程技术、农林医药和职业教育等学科。

2. 网易公开课 (<http://open.163.com/>) 是2010年网易公司推出的“全球名校视频公开课项目”,目前内容分为国际名校公开课、中国大学视频公开课、TED演讲、可汗学院教学影片、赏课、Coursera官方中文学习社区和公开课策划几个部分。网易还与爱课程(<http://www.icourses.cn/mooc/>)携手共建了“中国大学MOOC”(<http://study.163.com/mooc/find.htm#mcList>),提供来自于39所985高校的多学科优质课程。

3. 国内其他资源 学堂在线 (<http://www.xuetangx.com>) 是由清华大学研发的基于edX开放源代码的中文MOOC平台,目前由来自国内外著名高校的近200门课程供学习。

中国教育在线开放资源平台 (<http://www.oer.edu.cn>) 和新浪名校公开课都汇集了哈佛大学、耶鲁大学、斯坦福大学、麻省理工学院、复旦大学、浙江大学等国内外知名高校的开放课程,内容涉及哲学、法学、教育学、文学、理学、工学、医学、经济学、历史学、管理学等多学科。

本章小结

本章简要介绍了网络信息资源的含义和特点,详细介绍了 Google Scholar、SCIRUS 和 OJOSE 等学术搜索引擎,以及 Medical Matrix、Medscape、HON、Medical world search 等医学搜索引擎; HighWire Press、Biomed Central、PubMed Central、DOAJ、PLOS、Free Medical Journals、中国科技期刊开放获取平台、开放阅读期刊联盟、中国预印本服务体系、中国科技论文在线、中国学术会议在线等重要的开放存取资源;以及日渐为大众所熟悉的网络开放课程 MOOC。

(马 路 宋余庆)

思考题

1. 什么是搜索引擎?
2. 请简述搜索引擎的种类。
3. 试比较医学搜索引擎各自的特点。
4. 你平常习惯于使用哪个搜索引擎? 为什么使用它?
5. 目前医学搜索引擎越来越重视医学继续教育(CME)资源的建设, 你能列出 3 个有 CME 资源的网站吗?
6. 请设计一个课题利用搜索引擎查找相关信息。
7. 什么是开放存取资源?
8. 试在开放存取资源中找到 3 种儿科相关学术期刊。
9. 试从开放存取的资源中查找一本有关 H1N1 流感的图书。
10. 试从网络开放课程中查找信息素养相关课程。

第十一章 文献信息分析与利用

美国认知心理学家司马贺说:“人是通过搜索来解决问题的,所谓搜索就是提出策略并用其来解决面临的问题”。而思维就是主体对信息进行的能动操作,如采集、传递、存储、提取、删除、对比、筛选、判别、排列、分类、变相、转形、整合、表达等,搜索的目的是为了优化思维来解决问题。文献信息分析与利用的过程,也可以说就是优化思维解决问题的过程。在文献信息分析与利用过程中,对文献信息的洞察、分析和评估的过程,涉及思维肯定的判断所进行的可能为有形的或者无形的思维反应过程,并使科学的证据和日常的常识相一致,就构成批判性思维(critical thinking)。

本章重点阐述文献信息收集、整理、评价、分析与利用所涉及的基本概念、基本理论和方法。

第一节 个人文献信息的搜集与整理

一、个人文献信息的搜集

迄今为止,人类在优化思维解决问题方面大量使用的是信息方法,比如说良好的教育可以使人思维优化。文献信息分析与利用的本质就是将经过选择的信息(包括知识信息、自然信息、生活信息、实践信息)系统地输入用户的头脑,使其头脑中的意识结构变得更适于解决问题。扩大信息接触渠道,即持续性扩大用户的信息输入和贮存,使个体经常处于认知拓进状态,才能不断地在头脑中发生新的认知活动和取得新的认知结果,促进主体的思维能力不断得到锻炼而不断优化起来,也就是我们常说的终身学习。

(一) 广义的信息搜集渠道有

1. 视听法 是指人们通过阅读图书期刊等文献信息、收听广播、电视和录音的方式来扩大信息接触。在信息社会里,视听手段在扩大信息接触中所占的地位愈益显著。

2. 活动法 在原地原行业通过增加活动,也可以扩大信息接触。对于职场人士而言,可通过参加下述活动来扩大信息接触。①实践活动:即变革自然的生产实践活动、处理社会关系的实践活动和科学实验活动;②交往活动:如在学术界参加学术交流,课题论证会议,参观考察,技术鉴定会,迎来送往的社交活动,正式、非正式的交谈等,从中都可以得到不少信息。国外的研究结果证明,从会晤中获得的信息量要比看书高 70%,还有文献表明大约 1/3 的科技信息是通过非正式渠道传递,可见交往活动对于扩大信息接触的意义。

3. 移动法 人由一个地方移动到另一个地方,由一个职业领域移动到另一个职业领域,这是扩大信息接触的一个常用方法。如住院医生的轮转。

4. 工具法 指通过工具 A 来获取 B 信息的方法。例如通过检测头发中的所含微量元素而察知人体是否有饮食失调、少年多尿症和同新陈代谢有关的其他偏离正常状态的情况。下列两种方法也包括在工具法之中:

(1) 通过发明制造延长感官的器具与技术来扩大主体的信息接触。例如人的自然感官眼睛看不到病毒、分子、原子,但是人类发明创造了高功能的显微镜,使得眼睛能够看见了这些东

西,使人脑接触到了更多的病毒、分子和原子的信息;再如人体内部组织的信息,人们可以接触到的有限的,而当各种结构的内镜发明以后,体内组织的信息就更多更快地被看到了。

(2) 通过发明各种预先把信息汇集、存贮、组织起来,使人们在需要的时候能够迅速地将其搜集出来的方法与技术,大大提高了获取文献信息的效率,工具书、地图、网络数据库等都是大量信息的汇集体,人们从中获取所需的文献信息就方便了许多。

(二) 课题文献信息收集的原则

在收集课题文献信息时的信息接触,不是漫无边际的任何信息的接触。文献信息的收集总是沿着一定的认知目标或实践目标,也就是遵循下列的原则:①由近及远:即先查找课题最近几年的资料,再回溯到过去。因为有关方法、技术、理论性的资料,越近越有参考价值;②逐渐扩散:即先查有关专业或课题的核心资料,包括核心著作、期刊、资料汇编等,再逐渐扩大收集资料的范围,查找相关资料;③兼收并录:即不仅收集正面材料,也收集反面材料,不仅收集与自己的假设一致的材料,也要注意收集与自己的假设不同甚至相抵触、矛盾的材料;④注意收集原始文献,应尽可能不用第二次或多次转手文献,以避免文献信息利用出错。

二、个人文献信息的整理和积累

(一) 个人文献信息的整理

文献信息整理的方法主要包括对搜集的文献信息识别、筛选、摘录和组织编排等。整理阶段的主要任务是对获取的素材进行分门别类,也就是说,将素材搜集阶段所制作的资料卡、考察记录、调查表等按内容分成若干类,例如分成政策类、研究类和技术类等。

1. 政策类 政策类的素材可以细分为:①制定该项政策的原因和背景;②国家或地方政府对该问题制定的政策和法令;③党政领导人对该问题的讲话;④报纸杂志对政策的解释和评论;⑤政策的实施情况,实施后的效果,以及贯彻执行中遇到的问题;⑥其他国家对于类似问题所制定的政策和法令,或者所作的一些限制性规定。

2. 研究类 研究类素材可以细分为:①课题的概念和重要性;②研究的对象、范围以及研究的目的;③课题的发展情况和当前的研究水平;④该研究课题与其他领域的关系;⑤研究成果的可能应用前景。

3. 技术类 技术类素材可以细分为:①技术的兴起背景和开发目的;②技术的原理和技术的优缺点;③产品的结构、外形、包装和性能;④产品或技术的研制情况及其发展概况;⑤技术的推广应用情况,以及应用以后的社会和经济效果。

4. 人员、设备及经费类 此类素材可以细分为:①研究或者研制的方法、材料和设备;②国外对于同类研究或研制项目的投资情况;③国内外研究该课题的其他科研机构的情况,例如科研数量、人员素质、实验条件等;④国内外开发同一技术或研制同一产品的其他单位的情况,例如技术人员的数量和素质、生产能力、产品的市场占有率等。

在整理工作中,素材究竟分多少类、分哪些类以及如何归类,应当根据课题性质和所收集的素材内容而定。一般说来,分类越细,将来使用也就越方便。如申请国家自然科学基金项目,其分类可以参考基金申请书的要求划分为:①立项依据(包括研究意义、国内外研究现状及分析,需结合科学研究发展趋势来论述科学意义,或结合国民经济和社会发展中迫切需要解决的关键科技问题来论述其应用前景);②项目内容、研究目标,以及拟解决的关键科学问题;③拟采取的研究方案及可行性分析(包括有关方法、技术路线、实验手段、关键技术等说明);④本项目的特色与创新之处等。

分类以后,将卡片素材按类集中装入卡片盒内,可在每一类的前面设置一张导片,并在导片上端的突出部位标明这一类的类名。在个人计算机使用文献管理软件时代,可以按照分类细分文件夹。

分类对于批判性思维是很重要的,它将信息分成合适的小组,便于识别此信息与彼信息之间的联系。在批判性分析中,便于帮助读者作出正确的比较,做到“物以类聚”。

(二) 个人文献信息的积累

个人文献信息的积累是与搜集、整理同步进行的一项工作,只搜集不积累,就失去了搜集的意义。阶段性文献信息利用工作结束以后,对于这一研究课题的素材要妥善保管,以后还可以不断补充新资料。这些素材日后既可以用于类似的研究工作之中,还可能为个人提供新的研究课题思路。

在个人计算机普及前,传统文献积累的主要方法是做笔记、写卡片、全文复印、做批语或符号。其中做笔记有助于提高写作阅读能力,锻炼思考,培养揭示问题本质的能力和准确简练表达自己思想,即“不动笔墨不读书”。

阅读科研文章时,需要详细分析其研究成果或方法论对推进该领域的知识作出的具体贡献。如:该论文要旨到底是什么?论文的理由是否支持论辩?有支持的证据吗?论文中观点和我对该学科已经知道的相符吗?论文中观点符合别人对该学科的看法吗?论文中观点与我目前的研究相关吗,有用吗?论文对该学科之前的研究有什么贡献?论文里面有缺陷吗?

笔记不要不加区分地摘录信息,这样就会丢失批判的着眼点,找不出最相关的观点。笔记通常分成下面几种:

1. **提纲式笔记** 是将文章中所讨论的或者研究的主要问题,用提纲的形式记录在笔记本上。简单的提纲通常是论文大小标题的集合,详细的提纲则是在大小标题下面,用非常精练的文字写出这一部分的主要内容。

2. **摘要式笔记** 是将文章的内容取其精要,记录在笔记本上。摘要式笔记记载的内容通常有:文章讨论的基本问题,论证问题的事实或者理论依据,技术经济指标,重要公式,实验方法和实验结果等。

3. **引语式笔记** 是将文章中某些重要的文句原封不动地抄录在笔记本上。引语式笔记抄录的内容通常有:重要论点,新颖的观点,论证问题所用的新资料,比较精当、深刻的议论,文章中转引的某名人的名言等。引语式笔记所记录的内容,有相当一部分可能用到读者将来撰写的研究报告中,因此,抄录原文时,必须语句完整,文字准确,并且一定要注明出处,以便日后查对。

4. **心得式笔记** 是一种读后感式的笔记,它是读者阅读某篇文章以后,将自己的收获、体会、得到的启示以及产生的联想等记载在笔记本上。

课题的积累一定要尽量齐全,搜集到文献信息后,对文献信息予以阅读、标记、做笔记并加以科学组织的方法,一直是科学研究和个人文献组织和管理的经典方法。

(三) 个人文献信息处理软件

现代的文献信息搜集、整理、积累的方法是在继承传统的文献信息搜集、整理、积累的基础上,与计算机技术、电子文献和网络技术相结合发展而来,其搜集、整理、积累和管理文献的手段更加丰富多样,主要体现在电子全文阅览器、云电子笔记和个人文献管理软件三种软件上。

1. **电子阅览器的功能** 电子全文的阅读器主要有:CAJViewer、超星阅览器、书生阅读器、Apabi阅读器、PDF阅读器和Kindle等基本都具备书签、附注工具、文本编辑工具、高亮文本工具、标注工具、文本框工具、箭头工具、线条工具、矩形工具、椭圆形工具和铅笔工具等。方便读者在电子书刊上把重要的地方划线做记号或旁注供以后翻阅参考用。很多阅读器还支持全文检索。

2. **云电子笔记** 是一款跨平台简单快速的个人记事备忘工具。会议记录、日程安排、生活备忘,心得灵感都可快速记录到云笔记,支持拍照和添加图片作为笔记附件。支持安卓版,桌面版,网页版,手机网页版这几种形式。主要产品有云笔记、印象笔记和有道云笔记等。

主要功能有:①记录(收集和整理各种信息,并且可以使用多种设备进行跨平台阅读和分享;②撰写(利用其提供的编辑工具,创建一篇图文并茂的笔记);③同步(下载手机客户端并开



启同步功能,与云端自动同步保存)。

3. 个人文献管理软件 是个人用于搜集、记录、组织、调阅、自动引用文献的计算机程序。

(1) 主要功能:

1) 文献收集高效化:借助个人计算机、网络通讯和网络数据库,大量的二次文献数据通过个人文献管理软件直接检索和导入个人文献管理软件;大量的网页、PDF 等一次文献可以直接下载并识别出文献的标题、作者等字段,导入时自动根据文件内容创建题录建立数据库。支持不同二次文献数据的格式批量导入。

2) 文献管理智能化:整个文献的添加、删除、编辑、排序、去重等管理高度自动化,自动分组、统计分析、形成统计图表、可视化文本工具等智能化管理标志了文献管理已经初步具有智能化的特征。

3) 引文写作一体化:自动化办公软件和文献管理软件相关联,达到用户撰写文档时能直接从文献管理软件数据库中搜索到指定文献,并以欲投稿期刊格式插入文中指定位置,而无需手工输入,节省了时间,也减少了手工输入的错误率。另外,还提供多种期刊的全文模板,方便用户创作。

4) 资源管理多媒体化:个人文献管理软件突破传统文献资源管理载体的限制,成为可以有效管理文档、图片、电子资源以及视音频资料的个人多媒体资源中心,即个人多媒体图书馆。

5) 资料管理的云共享:网络版文献管理软件与个人单机版文献管理软件的无缝连接,二者之间既相互独立也可以相互数据同步更新,提供多人分享功能。把文献与文献、文献与活动、文献与人连接起来,在人际交流的互助互动过程中通过文献与知识共享,运用群体的智慧进行创新,以赢得竞争优势。

文献管理软件的功能贯穿于文献信息分析与利用的整个过程,即知识的积累获取、加工整合、交流共享与创新,大大提升了用户信息处理和文献信息消化吸收的效率。

(2) 常用的个人文献管理软件:

1) EndNote: 是美国 Thomson ResearchSoft 公司开发的文献管理软件,现为 Thomson Reuters 集团旗下产品。该软件为收费软件,有个人版和大客户版。在其主页(<http://www.endnote.com/>)有 30 天全功能试用版下载。它与 Reference Manager 和 ProCite 被认为是当今世界上最优秀的三个文献管理软件。2013 年 6 月推出最新版本 EndNote X7(具体操作见在线课堂)。

2) 医学文献王:是北京金叶天成(北京)科技有限公司开发的软件。全中文操作界面。2004 年推出第 1 版,最新版本为 4 版。国内第一款医学文献管理软件,该软件具有智能化文献检索、专业化文献管理、便捷化获取全文和自动插入参考文献的功能,主要用于组织、整合、管理不同来源的医学文献。<http://www.kingyee.com.cn> 可下载免费版,并提供收费专业版。

主要功能:①直接检索 PubMed、Cnki、万方、维普、谷歌学术等数据库,提供在线检索和浏览器一站式检索二种检索方式,尤其 pubmed 在线检索具有添加中英文主题词自动匹配检索的功能;②管理不同来源的文献资料;③通过医脉通积分即可获得全文传递服务;④自动生成参考文献和引用编号(具体操作见在线课堂)。

3) NoteExpress:是北京爱琴海乐之技术有限公司研制的软件。全中文操作界面,支持两大主流写作软件 WPS 和 WORD。2014 年 10 月推出最新版本为 3 版。<http://www.inoteexpress.com/CompanyWeb/download.do> 可下载个人免费版和集团免费版(具体操作见在线课堂)。

三、文献信息的评价

在文献信息整理过程中,用户总会自觉或不自觉地对原始信息价值作出评价。例如,在形式整理阶段要对明显重复的信息去重、对明显陈旧的信息剔除;在内容整理阶段要根据文献信息内容标记重点和文献利用等级等。评价的结果,不仅关系到原始文献信息是否有用,而且直

接影响到文献信息分析与预测成果的质量。文献信息价值评价环节并不仅贯穿于文献信息整理过程之中,而是存在于文献信息搜集环节和文献信息分析与预测产品的传递和利用环节,并对这些环节的工作质量产生影响。

(一) 传统文献信息价值评价标准

1. 可靠性 原始信息的可靠性一般包括四个方面的含义:①真实性:如所搜集的原始信息是深入实际的调查所得,还是道听途说或添油加醋的结果,有无夸大或缩小的现象等;②完整性:即所反映的问题是否全面,是否抓住了事物的本质;③科学性:主要是指信息的内容是否科学;④典型性:即原始信息是否具有代表性和典型意义。

评价文献信息可靠性取决于其所依附的文献的可靠性。一般可归结如下:

(1) 看作者:如知名专家、学者及其他科技人员撰写的文献一般比实业、商业,新闻界人士撰写的文献准确、可靠。

(2) 看出版机构:如著名高校、权威出版机构出版的文献可靠性较强。

(3) 看文献类型:如专利、标准文献比一般书刊可靠性大,产品说明书比产品广告可靠性强;期刊文章往往被视为声誉最好的信息源,因为文章在发表之前,要接受其他知名学者的审阅和挑选,即“同行评审”(peer review)。知名期刊的发表竞争是很激烈的,因此成功通过同行评审的文章往往是评价高的。

(4) 看来源:如官方来源的文献比私人来源的文献可靠,专业机构来源的文献比一般社会团体来源的文献可靠。

(5) 看被引用率:反复被他人引用的文献可靠性要强。

(6) 看引文:引用(或参考)的文献权威性越强。

(7) 看程度:如最终报告比进展报告可靠,正式标准比试行标准或标准草案可靠,分红配股方案比预案可靠。

(8) 看密级:机密信息可靠性强于公开信息,但弱于绝密信息。

(9) 看内容:文献本身论点鲜明、论据充分、数据翔实、逻辑结构严谨,则可靠性强。例如,阅读“研究方法”章节时,尤其要注意取样方法是否最合适,是否是有代表性的样本(representative sample),即考虑到了相关群体和情况可能出现的种类的样本。评价文献中样本规模和统计学意义。评价文献研究中是否有控制变量,例如是否使用对照组。注意文献中看法、事实、专家意见的区分以及与证据的关系。看法是被认为是真的,但没有注明或充分证据的观念。看法可能是一个人或很多人的观点。事实是通过经历、直接观察、测试或与证据比较来检查和证明的信息。专家意见是建立在专业知识上的,这些知识往往需要很长时间积累,或是建立在研究或直接经验之上。

(10) 看实践:已实际采用或被实践检验证明能达到预期目的的信息可靠性强。

(11) 看利益关系:作者和结果是否有利益关系。

2. 先进性 原始信息的先进性是指该信息所报道或反映的内容是否在某一领域原有基础上提出了新的理论、新的观点、新的假说、新的发现,或者对原有的理论、原理、方法或技术加以创造性地开发和利用。只要是报道或反映人们在生产过程、科学实验以及其他社会实践的新进展、新突破的信息,都可以认为具备先进性。先进性是一个相对的概念,是与原有的基础相比较而言的,通常可以用时间、空间、内容三个矢量合成的结果来衡量。

(1) 文献外部特征:①看文献类型:正在进行中的项目的试验小结、刚刚更新的数据库以及新近发表的专利文献、研究报告、上市公司文件、会议文献等所含信息的先进性强;②看出版机构:权威出版机构出版的文献水平较高,所含信息的先进性强;③看发表时间:最近发表的文献信息的先进性强。

(2) 文献计量学特征:根据文献数量的变化所反映出来的某一领域发展的阶段和水平以及



文献半衰期的变化所体现出来的文献信息老化规律,可判断信息是否先进。

(3) 文献内容特征:根据文献内容在理论上是否提出了新的观点、新的假说、新的发现,在应用上是否提出了新的原理、新的设计、新的方法或者开创了应用新领域等,判断信息是否先进。

(4) 信息发生源各地区或部门的基础条件、文化传统、宗教习惯、社会政治制度、生产力发展水平等方面的差异性,决定了反映这些地区或部门现实状况和水平的文献信息先进性也参差不齐,例如美国和瑞士的制药工业、德国 B 贝朗公司和日本 Kameka 公司的血脂分离系统、美国的网络技术和市场经济实践都是比较先进的,相应地,反映这些国家的相关文献信息一般也是先进的。

(5) 实践效果:根据文献信息对实践的贴近程度和超前水平以及文献信息使用后所产生的经济效益、社会效益和环境效益的大小可判断文献信息是否先进。

3. 适用性 所谓适用性,是指原始信息对于信息接收者可资利用的程度。一般来说,原始信息的适用性取决于课题和信息用户两大因素,如所选课题或拟选课题的背景、内容、难易程度、研究条件以及信息用户的信息吸收能力、条件、要求等。原始信息的适用性评价通常是在可靠性和先进性评价的基础上进行的,即可对可靠而先进的信息按照适用性的要求作进一步的筛选。

(1) 看信息发生源和信息吸收者吸收条件的相似性:任何信息都是在一定的条件下发生的。评价原始信息是否适用于文献信息分析与利用,可以看其是否具有与利用者相似的条件。一般来说,具备发生源或吸收者吸收条件相似的原始信息是适用的;反之,是不适用的。

(2) 看实践效果:一般来说,实践证明具有良好的经济效益、社会效益和环境效益的信息是适用的;反之是不适用的。

(3) 看战略需要:文献信息的评价它不仅解决当时、当地存在的问题,而且要服从社会发展的长远需要。因此,某一信息是否适用,除了要考虑其是否适应当时、当地需要外,还应当考虑其是否在未来适用。

(二) 网络信息资源的评价标准

由于网络信息发布自由,信息来源广泛,信息量极大,信息内容庞杂,质量参差不齐,给人们选择和使用信息带来了很大的不便。人们在选择利用网络信息时,必须对其进行评价以决定取舍。对网络信息评价可以根据用户的经验,但更科学的方法则是应当对网络信息的评价制定客观的评价标准。

对网络上的信息资源评价,比较典型的标准有两种:“10C”原则和 CARS 检验指标体系。

1. “10C”原则 所谓“10C”原则,是指用以下十项标准进行网上信息资源的评估:①内容(content):信息内容可以反映信息的本质。用户通过对信息的浏览,可以了解到信息的倾向性、目的性、学术性,再结合信息发布更新的日期,作出对信息内容的评价;②置信度(credibility):置信度指的是信息发布者的可信程度,亦即其发布的信息质量是否可靠。一般认为,权威机构或作者发布的信息可靠性要高一些,即置信度要高一些。用户可以从作者(机构)和作者的权威性、信息发布的目的以及网址的后缀名来判断信息的可靠性;③批判性思维(critical thinking):浏览信息时,搜索者应利用自己的知识和经验对信息内容进行判断,而不应盲目接受或一概摒弃;④版权(copyright):版权主要强调的是信息资源中所含的文字、图表、影像、声音等内容的所有权归属,用户在网上使用信息时,必须尊重版权;⑤引文(citation):一般来说,正式的或权威的网上信息会给出引用来源;⑥连贯性(continuity):用户从某站点选取信息时,必须考虑该站点是否可长期依赖,同时还必须考虑该站点的收费情况,用户从某站点搜索信息时,要注意信息内容的更新速度和成本的最佳结合;⑦审查制度(censorship):可根据站点对资源利用的审查制度和非公开领域的大小,了解某一站点或某一信息的可靠性和真实性,如,中国对提供互联网信息服务的经营性 ICP 实行许可证制度,网站的 ICP 核对;⑧可连接性(connectivity):某站点在面对多个用户的登录要求时,是根据什么机制来满足每个用户的要求的,能否在

较短的时间内进入,是否允许多种访问工具,内容是否对各种网络浏览器开放等;⑨可比性(comparability):信息在网上发布的同时,是否也以印刷品或CD形式发行,站点中的信息是否全面,用户是否需要针对不同时段的数据或统计数字进行比较等,这些都必须由用户结合自己的情况来考虑;⑩范围(context):在搜索信息之前,用户应首先确定自己的信息需求,然后确定使用什么样的资源来满足需求,同时又不会造成大量的数据冗余。最后再根据所获取信息重新调整自己的信息需求。

2. CARS 检验指标体系 CARS指的是以下几个指标:置信度(credibility)、准确(accuracy)、合理性(reasonableness)和支持度(support)。与“10C”原则相比,这一指标体系从纵向上更加深化、细化,但涵盖的广度上不及“10C”原则。

(1) 置信度:用户可用以下方法判断信息的权威性和可信度。

- 1) 作者的可信程度:作者或站点必须提供证据来保证信息的知识性、可靠性和真实性。
- 2) 作者的受教育程度:在某一领域内所受与该信息相关的培训与经历,寻找作者的头衔和职务,是否附有与作者联系的方式。
- 3) 站点的可靠性:站点的建立时间,在同类站点中的权威性以及数据更新速度等。
- 4) 通过措辞等信息表达方式以判断信息的可靠性,例如:匿名的作者或站点,语法错误或错字,否定性的论调等都会对置信度产生很大的影响。

(2) 准确性:准确性是指信息是否及时,是否确定,是否有细节,是否精确,是否综合全面。评价准确性,可从以下三个方面来考虑:

- 1) 及时性:经典著作不受时间限制,但许多网上信息却寿命极短,在搜索信息时,用户要注意信息发布的日期,是否已过时。
- 2) 综合性:任何结论或声称(隐含或不隐含)要发展一种完备的理论的信息源,都应该反映出完备性和精确性,换言之,信息应该具有综合性。
- 3) 适用性:进行信息评估时,重要的是确定信息是否为特定用户所需。
- (3) 合理性:指信息的公正性、客观性、节制性和前后一致性。

1) 公正性:指信息应提供公正合理的论证,而不是有选择、有倾向的论证,即使是竞争对手的观点和说法也应以正确的态度提供给用户。一个优良的信息源应该以一种客观的、理性的态度去论证和提供资料,而不应试图挑起用户的情绪,从而影响用户的清醒思考。

2) 客观性:完全的客观是不存在的,但一个好的信息源应不受个人的偏见和利益冲突的影响。对客观性最大的阻碍就是利益冲突,因为许多信息是靠使用户接受某些特定信息来获益的(通常是经济利益,但有时是政治利益,甚至是情感或心理利益),在这种情况下,用户应该非常谨慎地考虑该信息的客观性。

3) 节制性:是指当面对任何信息时应保持一颗平常心。首先根据自己的知识和经验判断信息的真实性,然后再用实践去检验其可信性,不要一看到不合常理的观点出现,就进行抵制,而应谨慎对其考察和判别。

4) 前后一致性:要求信息自身不自相矛盾。

(4) 支持度:支持度与信息的来源和确证有关,许多信息,特别是来自其他信息源的统计数据 and 事实数据,会增强信息的可信性。

1) 信息源:信息来自何方?其生产者利用了哪些信息源?有无参考书目或其他文献?作者是否提供了进一步联系的信息?该信息得到了怎样的支持?尤其要注意统计数字的引用及其出处。有些机构站点上提供有对产品、技术或生产过程的描述信息,如果用户自己可以分清事实和广告,那么这样的信息也是可靠的。

2) 可信性:主要指其他信源是否支持这条信息,如果某条信息确实论辩有力,置之有据,那么就会获得许多支持,至少会有人表示对该观点的赞同,所以用户必须找到能支持已获信息



的其他信源,以进一步确定信息的准确性和可靠性。

3) 外部一致性:是将新信息源中的信息和其他信息源中的信息进行对比。因为信息常常是新旧内容的交融,是已知和未知的结合,当信源中牵涉到用户已知的事实或论据时,外部一致性主要看它是否与过去的认识保持一致。如果信源对已知内容、信息源进行了歪曲,那么对未知内容也同样可能进行误导。

分类、筛选、区分、比较和对比都是批判性思维需要的基本技能,所以文献信息的识别、分析、评价、整理和积累这些要素是构成批判性思维训练的关键要素。

(周晓政)

第二节 文献信息分析

一、文献信息分析概述

(一) 文献信息分析的定义

文献信息分析主要是指以对大量已知文献信息的内容进行整理和科学抽象为主要特征的信息深加工活动。在此过程中,要对文献信息的价值进行评估,然后选取可靠的、先进的、实用的文献信息进行整序、统计分析,提取其中隐含的知识,从而获得增值的文献信息知识。

(二) 文献信息分析的意义与功能

1. 文献信息分析的意义 文献信息分析是利用文献信息的主要手段之一,学习文献信息分析具有非常重要意义。信息利用是信息管理的终极目标。所有的信息搜集、整理、存贮、检索等活动,其目的是为了供用户使用信息。而信息利用必不可少的前提就是要对信息进行分析。

(1) 从人才培养的宏观角度看:文献信息分析能力是大学生整体素养重要组成部分之一。正如本书绪论所介绍的“信息素养”是21世纪大学生必备的素养之一。作为信息素养的有机组成部分,在了解医学文献信息检索基本知识和基本过程,掌握各种医学文献数据库的检索技能之后,更应注重文献信息分析能力的培养。

(2) 从个人专业发展的微观角度看:对于从事医学科学研究和医疗实践的医学生而言,要有效地进行学习和科研,掌握文献信息分析技能是其未来生存和发展的基础。只有这样才能对文献信息的价值进行判断,迅速选取有用的文献信息,发现文献信息中隐含的相互关联,从而掌握相关学术领域知识,提高自己的学术水平。

(3) 从生存发展环境的客观角度看:在当今信息爆炸的时代里,文献数量呈指数方式增长,生物医学专业人员面临着海量的信息,如何利用文献信息分析技术,从大量的文献信息中获取自己所需要的那部分信息,日益成为生物医学专业人员需要掌握的重要技能之一。

2. 文献信息分析的功能 主要包括:

- (1) 整理功能:对信息进行收集、组织,使之由无序变为有序。
- (2) 评价功能:对信息价值进行评定,以达去粗(取精)、去伪(存真)、辨新、权重、评价、荐优之目的。
- (3) 预测功能:通过对已知信息内容的分析获取未知或未来信息。
- (4) 反馈功能:根据实际效果对评价和预测结论进行审议、修改和补充。

二、文献信息分析的步骤

文献信息分析的步骤

文献信息分析基本步骤可以分为两个部分,即传统的文献计量学的分析过程(步骤1~3)和在此基础上的共现聚类分析过程(步骤4~6)。

具体步骤如下(图 11-1):

1. 获取样本 首先确定欲分析的主题范围,在相应的文献数据库中检索到与该主题相关的文献记录,并下载相关文献记录。在文献信息分析中,一般要求文献记录达到一定的数量(经验上一般 3000~5000 条记录比较适宜),以避免随机因素造成的误差。对于文献量较大的主题,普通的计算机运算能力难以处理的,可以将文献记录分时间段(如每 5 年)、按照语种或者期刊的特征将文献记录分割处理,也可以用特定软件采取随机抽样的办法抽取样本。

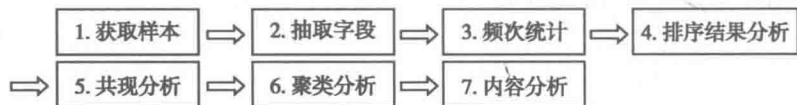


图 11-1 文献信息分析的基本步骤

2. 抽取字段 从已经下载的文献记录中抽取出将要分析的字段,比如期刊名称,作者姓名,主题词等,一般规范的文献数据库中这些字段都有固定的标识和位置,可以编制程序自动寻找和抽取所需的字段数据。

3. 频次统计 从下载的文献记录集合中抽取相关指标(如期刊、作者、引文等)进行频次统计,将统计指标按照出现频次由高到低排列,截取其中高于某个阈值的部分,如高频主题词、高频期刊、高产作者等,作为进一步分析的样本。

4. 排序结果分析 由于这些指标中出现频次较高的部分代表了该领域的重要事物,因此对于这些高频指标的分析可以表现出科研活动的结构。如某一领域的高频主题词代表了该领域研究人员发表论文中包含较多的主题概念,可以反映出该领域的研究热点;高频期刊可以反映发表该主题较多的期刊有哪些,甚至相关文献数量在不同年代的分布也可以反映该学科处于什么样的发展阶段。

5. 共现分析 通过两个条目同时出现的现象来发现条目之间存在着联系;条目间的共现频次高低,说明它们存在联系的程度,如论文著者的共现可以显示著者之间的科研合作关系,两篇论文如果在发表后经常被同行引用,说明它们的主题之间也有着密切的联系。这是共现分析的理论基础。

对于书目文献数据库中的某些字段,如果存在着两个以上的条目,对这些条目就可以进行共现分析。例如一篇论文可以有多个著者,因此就可以进行著者间的共现分析,即著者间的合著现象的分析;此外,论文的主题词、关键词、引文、乃至引文的作者、期刊等都可以进行共现分析。

6. 聚类分析 是指将物理或抽象对象的集合分组,利用统计分析软件,根据共现矩阵,采用聚类分析、社会网络分析等方法对相关指标进行聚类 and 表示,使其成为由类似的对象组成的多个类的分析过程。从数据的角度讲,聚类是通过计算分类对象在各个属性上的相似程度,将对象分类到不同的类或者簇的过程,使得同一个类中的对象有很大的相似性,不同类的对象间有很大的相异性。聚类与分类的主要区别在于聚类所要求划分的类预先未知的。

7. 内容分析 对通过共现聚类分析后得到的类别进行内容分析,以此说明该领域的科学研究活动的基本状况,如研究热点等。

下面通过实例说明文献信息分析的过程。例如,如果想要了解“当前胃癌治疗的研究热点是什么”,那么可以从 PubMed 数据库中检索到所有有关胃癌治疗的相关文献记录,抽取每篇论文的主题词,然后进行频次统计(表 11-1)。

这些主题词是所检索时间段内(本例为 2009 年 9 月 13 日检索的最近 5 年的胃癌治疗文献)出现频次最多的主题词,如胃肿瘤的手术治疗、胃肿瘤的药物治疗、联合化疗方案的应用、胃切除的方法、胃肿瘤的病理等,这些都是当前胃癌研究人员最为关注的主题,也是近年研究的热点。



表 11-1 胃癌 / 治疗高频主题词列表 (部分)

序号	关键字段	出现频次	百分比 %	累计百分比 %
1	Stomach Neoplasms / Surgery	1440	9.7813	9.7813
2	Stomach Neoplasms / Drug Therapy	1324	8.9933	18.7746
3	Antineoplastic Combined Chemotherapy Protocols / Therapeutic Use	651	4.4220	23.1966
4	Gastrectomy / Methods	366	2.4861	25.6827
5	Stomach Neoplasms / Pathology	354	2.4046	28.0872
6	Stomach Neoplasms / Therapy	328	2.2280	30.3152
7	Adenocarcinoma / Drug Therapy	321	2.1804	32.4956
8	Adenocarcinoma / Surgery	289	1.9630	34.4586
9	Gastrectomy	214	1.4536	35.9122

在此基础上,还可以进行主题词的共现分析。主题词的共现分析是根据主题词在同一篇论文中共同出现的次数来表示主题词之间的联系。一般认为,如果两个主题词频繁在同一篇论文中同时出现,往往表明这两个主题词之间具有比较密切的联系。

在 MEDLINE 数据库中存储的每一条记录都被标引 10 个左右的主题词以及相应的副主题词,尽管在同一篇文章中出现的主题词之间没有明确的关系,但是还是可以通过分析这些主题词出现的具体情形(如检查每个主题词的副主题词)而推理出这些主题词之间的关系。例如,如果 MEDLINE 数据库的一篇文献记录中有 x 疾病的病因和 Y 药的副作用这两个主题词,人们自然会想到“Y 药引起了 x 病”。通过查看原文就可以验证疾病和药物之间这种关系的有效性。因此,可以认为这些关系具有语义学上的意义。

对于代表研究热点的高频主题词统计它们共现次数后,可以形成一个高频主题词的共现矩阵,如表 11-2。其中的行和列的数字代表主题词,与表 11-1 中的主题词序号对应,如 1 代表的是 Stomach Neoplasms/Surgery; 矩阵中对角线的数字代表该主题词出现的总次数,非对角线上的数字表示的是两个主题词共同出现的次数,如 1 号与 2 号主题词共同在 60 篇论文中出现,1 号与 3 号主题词共同在 38 篇文献中出现。这些数据经过适当的计算后转换为相关系数,成为相关矩阵后,就可以显示出高频主题词之间的亲疏程度,为进一步分析做好了准备。

表 11-2 主题词共现矩阵 (局部)

	1	2	3	4	5	6	7	8	9
1	1440	60	38	351	178	6	13	289	150
2	60	1324	616	13	141	0	320	12	53
3	38	616	651	10	89	27	193	11	42
4	351	13	10	366	29	4	3	64	0
5	178	141	89	29	354	31	13	38	26
6	6	0	27	4	31	328	0	2	12
7	13	320	193	3	13	0	321	12	20
8	289	12	11	64	38	2	12	289	38
9	150	53	42	0	26	12	20	38	214

获得了相似矩阵之后,将其输入到统计分析软件(如 SAS, SPSS 等),选择相应的模块进行聚类分析(cluster analysis)。

以使用 SPSS 进行聚类分析为例,可以选择 Analysis-Classify-Hierarchical 过程,经过设置相应的参数后,对胃癌治疗的高频主题词共现矩阵进行分析,最后获得该研究领域高频主题词的

共现聚类分析树图(图 11-2)。

首先对聚类树图的结构进行分析。聚类树图中的最左边的一列标号(Label)和数字(Num)代表着高频主题词,由于采用的是系统聚类法的凝聚聚类算法,因此,最初每一个主题词都是单独的一个类,通过计算每一对主题词之间的相似性,首先是2号和3号主题词聚集成为一个类,然后它们又和7号主题词合成为一个类。图中最上方的带有数字的标尺表示分类对象之间的距离。随着被分类的对象(主题词)之间的聚类越来越大,距离最终所有的主题词都成为一个类,可以根据需要在不同的距离水平上分割整个聚类树图,通过树图的结构可以看到,所有的主题词大致可以分为三个部分:由2、3、7号词组成的一个类别(A),由1、4、8、5、9号主题词组成的一个类别(B),和由6号词单独组成的一个类别(C)。

对各个类别主题词之间语义关系的分析。基于凝聚聚类算法的原理,对聚类分析结果的语义分析也采用了“自下而上”的步骤。即首先获取各个小类的含义,然后把各个小类的含义组合成为大类的含义。具体而言,就是首先从每个小类中关系最近的两个主题词着手,分析二者之间的语义关系,获得该类的“种子”概念,在“种子”概念的基础上,根据同类别中其他主题词与该“种子”的距离,逐次加入主题词,丰富该类别的内容,一般而言,距离比较远的主题词往往是该核心的相关因素,如具体的应用或者影响因素。本例中,对于3个高频主题词的类别中的主题词进行具体的语义分析,可以发现:

在A类中,“Stomach Neoplasms/Drug Therapy, 胃肿瘤/药物治疗”(2)与“Antineoplastic Combined Chemotherapy Protocols/Therapeutic Use, 抗肿瘤联合化疗方案/治疗应用”(3)组合在一起表明的是对胃肿瘤采用联合化疗,加上“Adenocarcinoma/Drug Therapy, 腺癌/药物治疗”(7)表明这一类主要是关于胃腺癌的联合化疗的主题。

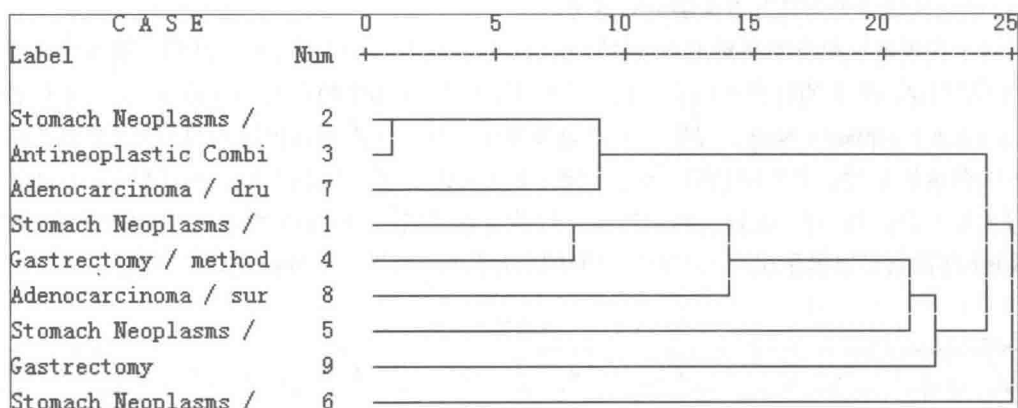


图 11-2 胃癌治疗高频主题词聚类分析结果(部分)

在B类中,“Stomach Neoplasms/Surgery, 胃肿瘤/手术”(1)与“Gastrectomy/Methods, 胃切除/方法”(4)表明是关于胃肿瘤外科手术切除治疗的主体,同样,“Adenocarcinoma/Surgery, 腺癌/手术”(8)也是说明主要是针对胃腺癌的手术切除治疗,“Stomach Neoplasms/Pathology, 胃肿瘤/病理”则是说明这种疗法会受到不同的病理类型的影响。这一类别最边缘的主题词“Gastrectomy, 胃切除”(9)没有副主题词,则是表示对胃切除的概述,是对上述主题的强化,由于与其他主题词距离比较大,也可以忽略不计。

C类则是单独一个单词,“Stomach Neoplasms/Therapy, 胃肿瘤/治疗”(6)则表示的是对各种治疗方法的概述性研究主题。

由此可以总结出目前胃癌治疗研究的主要方向有三个:①胃癌胃腺癌的联合化疗;②胃腺癌的手术切除治疗(包括病理因素的影响);③胃肿瘤各种治疗方法。

通过对特定领域或者学科的高频主题词的共现聚类分析,客观地反映出当前该领域研究的



热点。其他表现文献内容的标识(如关键词、分类号)等,也可以用于此类分析。此外,一些表现文献外部特征的标识,如作者、引文等,也可以进行共现分析,如作者的合著分析、引文的同被引分析、作者的同被引分析,这些为展示某一学科领域里科学研究获得的结构和特点提供手段。这些分析方法都是基于共现的聚类分析,其原理大同小异。

三、文献信息分析的方法

目前,文献信息分析方法从文献计量学方法(bibliometrics),逐渐向文本挖掘(text mining)和基于文献的知识发现(literature based discovery)纵深发展。其中,文献计量学就是借助文献的各种特征的数量,采用数学与统计学方法来描述、评价和预测科学技术的现状与发展趋势的学科。文献计量学的方法可以分为书目计量法和引文分析法。

(一) 书目计量法

书目计量法分析的指标主要是文献的书目信息,如文献量(各种出版物,尤以期刊论文和引文居多)、作者数(个人、集体或团体)、词汇数(各种文献标识,如主题词、关键词和分类号等)。统计这些特征出现的规律,就可以显示出该领域或者主题的研究活动的特点。例如通过对检索到的“胃癌治疗”的文献的外部特征分析,可以了解到该研究领域的发表相关论文较多的核心期刊,该领域的重要作者等。对于这些指标的分析研究形成了文献计量学的重要定律,这些定律构成了文献计量学的主要内容,对这些定律应用也是文献信息分析的主要手段。

1. 布拉德福定律 关于某一特定课题、学科或领域的论文,我们称之为该课题的“相关论文”。相关论文在期刊中不是均匀分布的,即除了发表在本专业的期刊上,也发表在相关专业的期刊上。例如关于胃癌治疗的论文,大多数都发表在少数几种肿瘤临床类专科的期刊上,如《中华肿瘤杂志》《癌症》《中国肿瘤临床》等;但也有相当数量的论文发表在大量的相关专业的期刊上,如《中华病理学杂志》《中华放射肿瘤学杂志》《中华护理杂志》等。这种现象被称作相关论文在期刊上分布的集中-离散。某个专业里发表相关论文较多的期刊被称作该领域的核心期刊,目前,核心期刊已经成为评价论文学术水平的替代指标。

最早注意到这个现象并对其进行定量研究的是英国文献学家布拉德(Samuel Clement Bradford)。1934年,布拉德福选择了应用地球物理学和润滑领域为样本的主题,统计了1929—1931年间490种期刊,1727篇论文(表11-3)。对这些期刊按照发表相关论文的数量多少从高到低排列起来,如果按照期刊年平均发文量在4篇以上,4篇以下、1篇以上,和1篇及不足1篇的标准,将这些期刊分为三个区域的话,结果发现:这三个区域中的相关论文数大致相等,而相继各区的期刊数基本上成等比数列(接近5)。

表 11-3 布拉德福统计的两个学科的文獻数据

分区	期刊载文数量	应用地球物理学		润滑	
		期刊量	论文量	期刊量	论文量
1	>4	9	429	8	110
2	1~4	59	499	29	130
3	1	258	404	127	152

这一发现被后人称为布拉德福定律,布拉德福定律的文字描述为:如果将科学期刊按其刊载某个学科领域的论文数量以递减顺序排列起来,就可以在所有这些期刊中区分出载文量最多的“核心”区和包含着与核心区同等数量论文的随后几个区,这时核心区 and 后继各区中所含的期刊数成 $1:a:a^2\cdots$ 的关系($a>1$)。

布氏定律主要反映的是同一学科专业的期刊论文在相关的期刊信息源中的不平衡分布规律。这种规律主要是应用于图书馆和期刊文献情报学研究领域,如指导文献情报工作和科学评

价等,以切实提高文献信息服务和信息利用的效率和科学评价的科学性。

对于信息用户而言,布氏定律可以帮助确定感兴趣的研究领域的核心期刊,寻找到重要的信息源。比如,可以利用布拉德福采用的经典方法,从权威文献数据库(如 PubMed)中下载感兴趣的主题的相关文献记录,对这些记录进行简单的转换(如用 Microsoft Excel)后,抽取出来源期刊字段加以统计,就可以了解到发表某领域或主题相关论文较多的期刊,以此作为今后定期浏览关注的核心期刊。

在对期刊分布的研究中,和信息用户有关的还有关于核心期刊概念的界定。从布氏定律的文字描述中了解到,最早提出的核心期刊就是指按照发表相关论文数目多少等级排序期刊表中,位于核心区的那部分期刊。但是,在目前,核心期刊的定义已经转化为“某学科所涉及的期刊中,刊载论文较多的(信息量较大的)、论文学术水平较高的,并能反映本学科最新研究成果及本学科前沿研究状况和发展趋势的,较受该学科读者重视的期刊”。也就是说,核心期刊成为评价期刊质量,进而评价论文质量的指标。尤其是近 10 多年来,核心期刊论文的发表数量已经成为国内各高校、研究机构、企事业单位人事、科研等机构部门评定科研人员科研业绩的“硬指标”。目前,我国在自然科学领域中,比较重要的核心期刊目录包括北京大学图书馆的《中文核心期刊要目总览》,科技部中国科技信息研究所的《中国科技期刊引证报告》和中国科学院文献情报中心负责研制和发布的《中国科学引文数据库来源期刊》。可以参考这些核心期刊目录了解某个领域中的核心期刊有哪些。

2. 洛特卡定律 从文献的角度出发,作者是一篇论文外部特征之一,尽管有些文献的作者可能是匿名的,但是在文献数据库中作者这一字段仍然是必不可少的。目前,随着科学研究合作的倾向越来越紧密,一篇论文的作者数量也呈现出不断增加的趋势。

从作者的角度上说,研究者在学术上表现出的能力和工作效果,通常用其生产的学术文献的数量来衡量,因此,一个作者发表论文的数量往往被当作衡量其科学生产率的一项指标。一般而言,专家们认为:对于某一个学科领域而言,大多数的作者发表了 1~2 篇的论文,只有少数人发表了大量的专业论文,人们发现某一时期、某一主题内作者发表论文的数目与相应的作者数目之间也有存在着规律。例如 1997 年,有学者应用有关作者发表文献的定律,确定了我国地震系统的高层次人才数约为 513 人,可以作为当时该学科专业的研究员的人数,并且与实际情况基本符合。

最早运用数学公式定量地表达这个规律的就是洛特卡定律。1926 年,在保险公司供职的洛特卡(Alfred James Lotka)分别对化学领域和物理领域的作者发表论文数目进行了统计,他所抽取的样本中,发表了 1 篇论文的作者化学领域里有 3991 人,物理领域里有 784 人;发表了 2 篇论文的作者分别是 1059 和 204 人……,最多的在化学领域里有 1 人发表了 346 篇论文,物理领域有 1 人发表了 37 篇论文。与布拉德福对核心期刊的分析一样,洛特卡也将作者名字按照发表论文数目的多寡由高到低排列,并分别赋予等级排列的序号。经过直线拟合等方法,最后得出如下数学公式:

$$f(x) = \frac{C}{x^n}$$

其中, $f(x)$ 为写 x 篇论文的作者占作者总数的百分比, x 为论文篇数, C 为某主题领域的特征常数。

洛特卡推论认为:“检验发现,生产 2 篇文章的作者大约是生产一篇文章作者数的 1/4,生产 3 篇文章的作者数是生产 1 篇文章作者数的 1/9……,写一篇文章的作者数是全体作者的 60% 左右。”人们称之为平方反比定律。

受洛特卡定律的启示,普赖斯(Derek John De Solla Price)提出了核心生产者的“平方根定律”,即在某一特定领域中,全部论文的半数系由该领域中全部作者的平方根的那些人撰写的。

这些人无疑是核心或高产作者。

洛特卡定律作为描述文献作者分布的一个计量学的经验规律,可以表述某一时期内、某一主题领域内,文献作者的分布情况,这对了解感兴趣的领域内科学研究活动的规模和核心作者十分重要。从人类学的角度来理解洛特卡定律,它可以表现为“在人类发展过程中,具有不同个性的个体之间对人类进步所作出贡献的差异”。

3. 齐夫定律 无论中文还是英文,构成论文的最基本要素就是词汇。最初,是从事速记和电话通讯的人们比较关注单词的使用频率问题,这个问题继而泛化成为“人类在使用自然语言的时候是否也有规律可循”的问题。这个研究领域中最著名的成果就是齐夫定律。

齐夫(George Kingsley Zipf)是美国哈佛大学的语言学家和心理学家。他在1948年出版了《人类行为与最省力法则 - 人类生态学引论》,提出了著名的“最省力法则”(Principle of Least Effort)。最省力也译作最小努力,是指人类在行为中,例如从A地到B地时,可以走不同的路;从经济上、安全上、时间上并结合本人的主观条件(如身体情况)及客观情况(所处的环境)等种种因素来考虑,设法选择一条最符合自己要求的道路,使得自己付出的“力”最小。

为了说明人类在利用语言表达思想时也在遵循最省力法则,擅长语言和心理学的齐夫用大量的统计数据来验证前人有关词频分布规律的研究成果,作为一个最省力法则的实际例子。他统计了在一定长度的文章中每一个词出现的频次,将这些词按照其出现频次从高到低排序,可以得到词的序号,结果发现词的频次和序号数量关系特征呈双曲线分布,即:

$$fr \times r = C$$

其中 r 为单词的等级序号, fr 为 r 所对应的频次, C 为常数。

从最省力法则来讲,任何语言中,凡是使用频率最高的词,功能总是不会太大;因为其本身在这个场合中价值小,因而传递它们所需要的“力”就不大。人们在表达思想的时候,会受到两个方向相反的力的作用,一种是单一化力,希望尽量简短;另一种则是多样化力,希望能被对方理解,单一化力与多样化力取得平衡,使自然语言词汇的使用频次分布呈现了双曲线。

在信息检索领域里,齐夫定律被广泛应用于叙词表的编制和自动标引的研究中,以提高计算机检索的效率。例如在一个检索数据库中,如何运用计算机自动给一篇文献的赋予相应的关键词?一般可以用计算机提取出原始文献中的专业术语,统计其频率,研究分布特征,根据这些词的使用频率的词最后决定该篇论文的关键词。例如卢恩(Luhn)的频率自动索引方法就是统计词频之后,去掉高频词,去掉后缀,找出相应的词干最后确定表现论文内容的区分力(resolving power)最高的关键词(图11-3)。

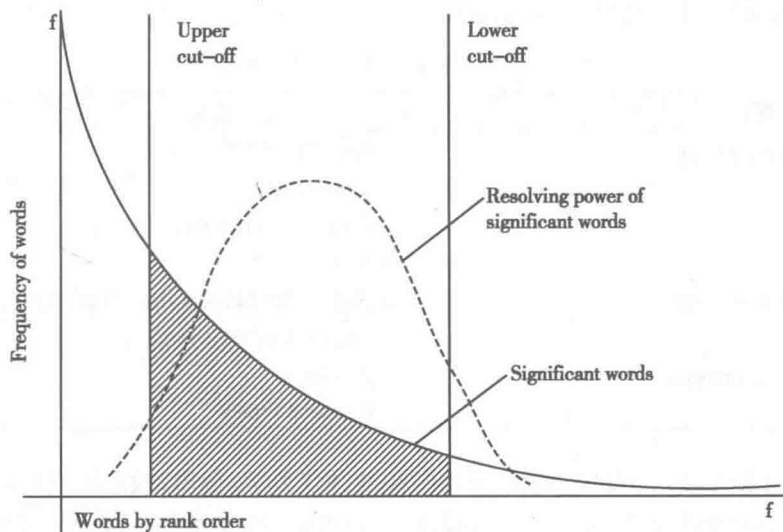


图 11-3 卢恩的词频自动标引原理

4. 普赖斯指数 文献是存在着“老化”的现象。这是因为有些文献中所含的情报已失效,或者文献信息已包含在其他文献之中,原有的文献被更新的文献所代替,或者人们对某一个课题的研究兴趣下降造成有关文献的利用减少,或者是由于新观念与新理论的发现,使得文献发生了老化(obsolescence, aging),因此,文献老化就是指科学文献随其“年龄”的增长,其内容日益变得陈旧过时,作为情报源的价值不断减小,甚至完全失去利用价值的现象。

对于文献老化有多种度量指标,如半衰期(half life)、普赖斯指数(price index)和期刊剩余有益性指标等。其中,普赖斯指数因其计算简单使用方便而得到广泛的应用。

普赖斯指数由 D. S. Price 于 1971 年提出,普赖斯发现文献的作者在引用文章时倾向于引用最近发表的论文,因而使得较早发表的文献的利用率逐渐降低。由此,普赖斯将被引用的文献分成二类,一类是经典文献,即年龄超过 5 年仍被引用的文献,另一类是则是过渡文献,即年龄不大于 5 年的被引文献。普赖斯认为,引文中,经典文献的数量与过渡文献的数量之比是文献老化的重要的特征之一。因此,普赖斯指数就是某一知识领域内,把对年限不超过 5 年的文献引文数量与引文总数之比。用该指数可以度量文献的老化速度和程度。普赖斯指数愈大,半衰期则愈短,文献老化的速度则愈快。

普赖斯指数主要用于科学和科技史研究中,比较不同学科的发展的速度,揭示科学发展的规律,反映出人类如何继承和发展科学知识。也可以用于评价文献。

(二) 引文分析方法

引文除了可以用于检索,还可以通过一篇论文被其他文献引用次数来表示该论文的影响力或者重要性,这一类的分析叫做引文分析(citation analysis)。引文分析法是文献计量学分析中的经典方法,也是应用最为广泛的方法之一。

1. 引文分析的定义 利用各种数学及统计学的方法进行比较、归纳、抽象、概括等的逻辑方法,对科学期刊、论文、著者等分析对象的引用和被引用现象进行分析,以揭示其数量特征和内在规律的一种信息计量研究方法。

2. 引文分析的指标 美国费城科学情报所(Institute for Scientific Information, ISI)是科学引文索引的创始者,2001 年,他们又根据引文分析的原理,推出一种定量地评估科学研究水平的重要研究工具:基本科学指标(Essential Science Indicators, ESI)。它是 ISI 网络集成服务平台的一个重要组成部分,网址为 <http://esi.isiknowledge.com>,用户也可以通过 Web of Knowledge 网站平台使用这个工具。

ESI 的内容包括引文排位(Citation Rankings)、高被引论文(Most Cited Paper)以及引文分析(Citation Analysis)三大主要模块(表 11-4)。

表 11-4 ESI 的 3 个主要模块及其功能

模块	功能
Citation Rankings	- Scientists - Institutions - Countries/Territories - Journals
Most Cited Papers	- Highly Cited Papers (last 10 years) - Hot Papers (last 2 years)
Citation Analysis	- Baselines - Research Fronts

(1) 引文排位模块:可以按照领域查询作者、机构(大学、公司、政府研究机构)、国家和期刊在一定时间段内的发文数、被引次数和排位,也可以根据具体的作者、机构、国家和期刊检索它们在各领域的排位情况,一般是按照它们总被引频次,按照一定的阈值或者百分比截取高被



引的作者、机构、国家和期刊。

(2) 高被引论文模块：包括高被引论文和热点论文列表。

1) 高被引论文：是 ESI 根据论文在最近 10 年内的被引频次，选择靠前的 1% 范围内的论文形成高被引论文列表。可以按照学科领域、作者名、机构名、期刊名、国家名来浏览这些论文，也可以输入论文的具体线索进行检索。

2) 热点论文：指与相同领域和相同出版年的其他论文相比，出版后很快就有高被引频次的论文。这些论文因为在很短的一段时间里得到关注，因此称之为热点论文。选择热点论文要求其出版年龄不能超过 2 年，而且是在当前的 2 个月里被引，按照相应的条件，选取其中前 0.1% 的论文。

(3) 引文分析模块：包括基线 (baseline) 和研究前沿列表。

基线由平均被引频次和百分点两个表格组成。

平均被引频次见表 11-5，表显示的是最近的 10 年间各个领域论文的平均被引次数，表格中的数字是“引文数 / 论文数”，例如学科领域 (第一列) 中的临床医学与年代 (第一行) 中的 2000 对应的数字为 20.30，表示 2000 年发表的临床医学论文到目前为止平均被引了 20.30 次。另外，最后一列 “All Years” 列出的是各领域 (或者全部领域) 的 10 年平均值，可以作为科学家、机构、国家以及期刊排位表给出的单篇被引值的基线。

表 11-5 ESI 中的平均被引频次 (局部)

Fields	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	All Years
All Fields	17.96	17.35	16.13	14.72	12.84	11.06	8.69	6.06	3.71	1.30	0.17	9.84
Agricultural Science	11.86	11.93	10.85	9.93	9.04	7.67	5.84	4.16	2.36	0.71	0.10	6.24
Biology & Biochemistry	28.83	28.08	25.61	23.03	20.20	17.00	13.03	9.98	5.45	1.96	0.21	16.04
Chemistry	16.42	16.52	15.06	14.63	12.90	11.35	9.28	6.62	4.15	1.59	0.17	9.83
Clinical Medicine	20.98	20.30	19.12	17.74	15.91	13.72	11.00	7.65	4.59	1.52	0.20	11.89
Computer Science	6.64	6.14	6.42	6.49	4.17	3.00	2.36	1.41	1.36	0.45	0.07	3.14
Economics & Business	10.41	9.77	8.71	8.56	7.24	6.02	4.40	2.80	1.53	0.49	0.10	5.20

百分点列出的是各学科论文按照被引频次由高到低排列后，累计被引频次的特定的几个百分位点所对应的被引频次。如果将各个学科论文的被引用次数的按照由高到低的排列，并统计出各个被引次数的频次，如被引 100 次的有 80 篇，99 次的有 98 篇……，这显然将是一个很大的表格，为此，ESI 采用了 6 个百分位点：0.01%，0.10%，1.00%，10.00%，20.00% 和 50.00%，列出累计被引频次达到前 0.01% 等位点所必需的最少被引用次数。例如，表 11-6 中农业科学的 0.01% 行与 1999 年交叉的表格中列出的数字为 494，表明 1999 年发表的农业论文只有被引用了 494 次才能进入前 0.01%。如果了解了某一篇文章的被引次数，通过这个表格可以知道该论文在本领域所有论文被引次数排序中的位次。

研究前沿实际上是一组高被引论文，也称之为核心论文 (core papers)，根据这些高被引论文两两同时被后来论文引用的次数，按照特定的阈值把所有具有密切关系的论文聚成一类。统计同一个类别的论文中出现频次较高的单词和短语，以此来表示该类别的主题。

在 ESI 检索平台中，可以通过选定学科来浏览各学科的前沿，也可以输入查询词获取包含该词的前沿类别。例如临床医学的研究前沿如图 11-4 所示，对每个类别的描述包括了该类别论文的主题 (如晚期非小细胞肺癌、体细胞表皮生长因子受体突变等)、高被引论文的数目 (反映基础文献的规模)，被引总次数 (反映研究前沿的规模)，平均被引次数 (反映集中的程度) 以及论文的平均年龄 (反映的该类别的现实性或者热度)。可以查看相应类别的高被引论文。

表 11-6 ESI 被引频次的百分点表

All Fields	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	All Years
0.01%	1233	1245	1170	1057	845	664	524	344	224	90	21	820
0.10%	490	462	421	370	318	258	204	139	85	36	8	301
1.00%	164	155	142	127	108	92	71	50	31	13	3	99
10.00%	42	41	38	35	30	26	21	15	9	4	1	24
20.00%	24	24	22	20	18	16	13	9	6	2	1	13
50.00%	8	8	7	7	6	6	5	3	2	1	0	4

Agricultural Sciences	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	All Years
0.01%	494	552	309	416	300	343	247	73	53	27	84	301
0.10%	221	220	191	154	155	111	81	49	31	14	8	134
1.00%	94	88	77	68	62	50	36	25	15	7	2	55
10.00%	29	29	27	24	22	19	15	11	6	3	1	17
20.00%	18	19	17	16	14	12	10	7	4	2	0	10
50.00%	6	7	6	6	5	5	4	3	2	1	0	3

RESEARCH FRONTS RANKINGS IN CLINICAL MEDICINE

Sorted by: Citations

1 - 20 (of 1438)

Page 1 of 72

View	Fronts	Papers	Citations	Citations Per Paper	Mean Year
1	 ADVANCED NON-SMALL-CELL LUNG CANCER: SOMATIC EGFR MUTATIONS; NON-SMALL-CELL LUNG CANCER PATIENTS TREATED; EPIDERMAL GROWTH FACTOR RECEPTOR GENE MUTATIONS; NON-SMALL CELL LUNG CANCER PATIENTS TREATED	48	19,186	399.71	2005.2
2	 WORLD HEALTH ORGANIZATION METABOLIC SYNDROME; METABOLIC SYNDROME USING TWO PROPOSED DEFINITIONS; METABOLIC SYNDROME IMPROVE IDENTIFICATION; METABOLIC SYNDROME DEFINED; 10-YEAR CARDIOVASCULAR DISEASE RISK	38	9,698	255.21	2004.4
3	 HUMAN CARDIAC STEM CELLS; MULTIPOTENT CARDIAC STEM CELLS; ADULT CARDIAC STEM CELLS; CARDIAC STEM CELLS POSSESS GROWTH FACTOR-RECEPTOR SYSTEMS; CARDIAC STEM CELLS DELIVERED INTRAVASCULARLY TRAVERSE	44	9,496	215.82	2005.1
4	 MICRORNA EXPRESSION PROFILES CLASSIFY HUMAN CANCERS; MICRORNA GENE EXPRESSION DEREGLATION; MICRORNA GENE EXPRESSION PROFILE, HUMAN MICRORNA GENES; MICRORNA MIR-155 EXPRESSION	44	8,968	203.82	2006.2
5	 CARDIAC RESYNCHRONIZATION THERAPY USING LEFT VENTRICULAR PACING; CHRONIC HEART FAILURE RECEIVING CARDIAC RESYNCHRONIZATION THERAPY; HEART FAILURE [THE CARDIAC RESYNCHRONIZATION-HEART FAILURE (CARE-HF) TRIAL EXTENSION PHASE]	43	7,723	179.60	2005.4

图 11-4 2009 年临床医学研究前沿列表

3. 引文分析的工具 利用引文分析的原理,人们开发出很多工具,主要用于表示某个学科领域的研究历史和进展。

(1) HistCit: 由《科学引文索引》的创始人尤金·加菲尔德(E. Garfield)等于 2001 年推出,是一个引文分析可视化系统。其主要功能是可以将某个领域的高被引论文按照发表时间先后顺序自动生成引文时序网络图,其网址为 www.histcite.com,可下载 30 天免费试用版。

该系统针对从 ISI Web of Science 上检索并下载的某一专题方面的文献记录(包括其引文信息)进行处理,将高于一定阈值的引文列出,按照被引论文发表年份顺序,绘制出网络图(图 11-5)。引文时序网络图展示某个研究主题的论文源流、最初著者以及该研究主题发展的来龙去脉,迅速对某一专题研究历史进行回顾,指导用户的研究工作。

此外,该软件可以把检索到的来源文献按照期刊名称、著者、年份分别进行排序。对于被引文献,还按照被引频次的高低进行排序,同时,点击引文时序网络图中的结点,可以看到引用该文献的所有文献以及被该文献引用的所有文献。

加菲尔德等人应用这一软件已经先后对有关数字图书馆、情报科学、信息计量学、共焦显微镜、实验胚胎学、基因组测序等专题研究以及某一段时期有关某一专题研究的杂志的发文情况进行了引文时序分析,分别生成了引文编年图。

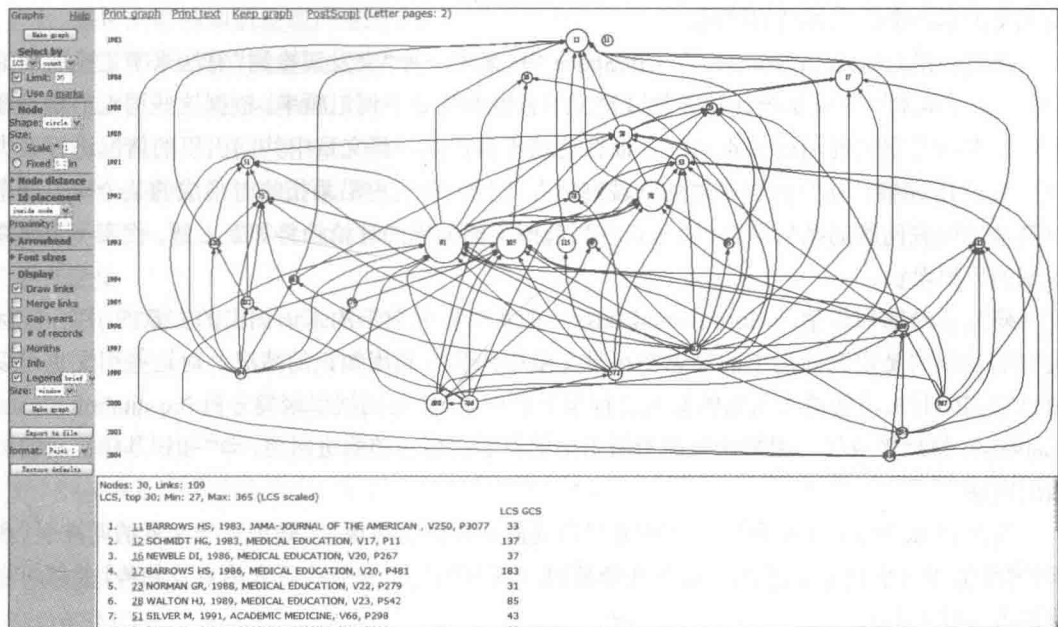


图 11-5 “以问题为基础的学习(PBL)”研究的引文时序网络

(2) Bibexcel: 由瑞典的于默奥大学(Umeå University)社会学系的 Olle Persson 开发, 可以在 www.umu.se/inforsk 网站上免费下载。

Bibexcel 是一个文献计量分析工具包。可以分析来从 Web of Science 等书目数据库中下载的记录, 该软件可以从下载的文件中抽取用户指定字段, 如标题中的词、作者、期刊、引文、被引作者、被引期刊等, 然后统计其出现的频次, 进行共现统计, 最后生成共现矩阵。其结果可以输入到其他软件(如 Pajek, Excel, SPSS 等)做进一步分析。此外, 用户还可以自定义输入记录的字段标识和分隔符, 具有更大的灵活性。

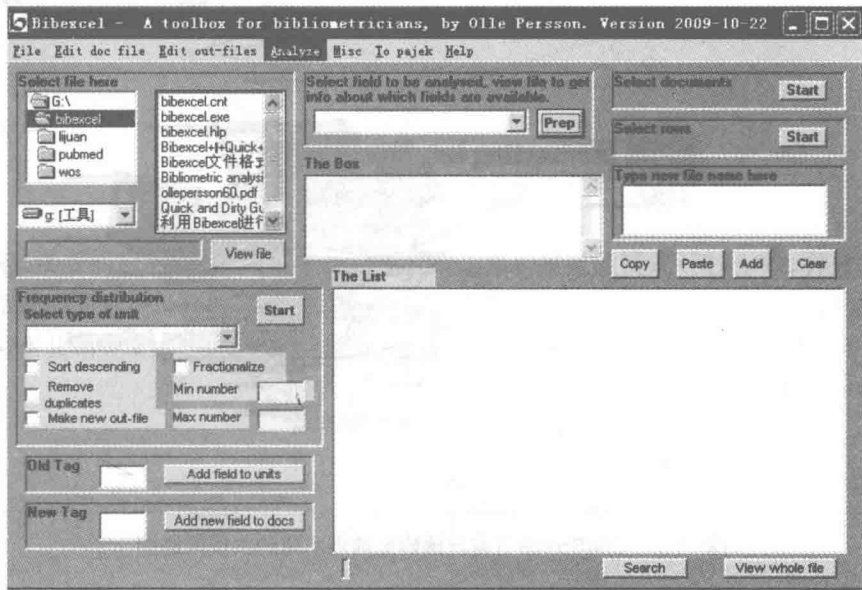


图 11-6 Bibexcel 的输入界面

(3) CiteSpace: 由美国德雷赛尔大学信息科学与技术学院的陈超美开发。该程序可以登录到 cluster.cis.drexel.edu/~cchen/citespace 后免费下载。

CiteSpace 的开发陈超美认为: 可以用“研究前沿”和“知识基础”随着时间相对地变化

情况来表示一个研究领域的状况。

研究前沿 (research fronts): 在 CiteSpace 中, 采用一种“突发词检测”算法来确定研究前沿中的概念, 基本原理就是统计相关领域论文的标题和摘要中词汇频率, 根据这些词汇的增长率来确定哪些是研究前沿的热点词汇。根据这些术语在同一篇文章中共同出现的情况进行聚类分析后, 可以得到“研究前沿术语的共现网络”。因此, 研究前沿系指临时形成的某个研究课题及其基础研究问题的概念组合, 也是正在兴起或突然涌现的理论趋势和新主题, 代表一个研究领域的思想现状。

研究前沿的知识基础 (intellectual base): 即含有研究前沿的术语词汇的文章的引文, 实际上它们反映的是研究前沿中的概念在科学文献中的吸收利用知识的情况。对这些引文也可以通过它们同时被其他论文引用的情况进行聚类分析, 这就是同被引聚类分析 (co-citation cluster analysis), 最后形成了一组被研究前沿所引用的科学出版物的演进网络, 即“知识基础文章的同被引网络”。

因此, CiteSpace 就是利用三个网络 (“研究前沿术语的共现”, “知识基础文章的同被引”和 “研究前沿术语引用知识基础文章”) 在随着时间演变的情况来寻找研究热点及趋势, 并以可视化的方式展示出来。

具体做法包括: 首先, 下载相关文献, 用户可以从 Web of Science 中检索并以固定格式下载某一主题的文献记录, 主要包括作者、题目、摘要和文献的引文等字段, 将检索到的文献记录输入到系统之后, 系统会生成对文章标题、文摘和描述词的频率统计, 然后设置参数, 如确定要分析领域的总的时间段范围和分割后每一个时间片段的长度等; 运行系统后会得到“研究前沿术语的共现网络”(图 11-7), “知识基础文章的同被引网络”和“研究前沿术语引用知识基础文章网络”。

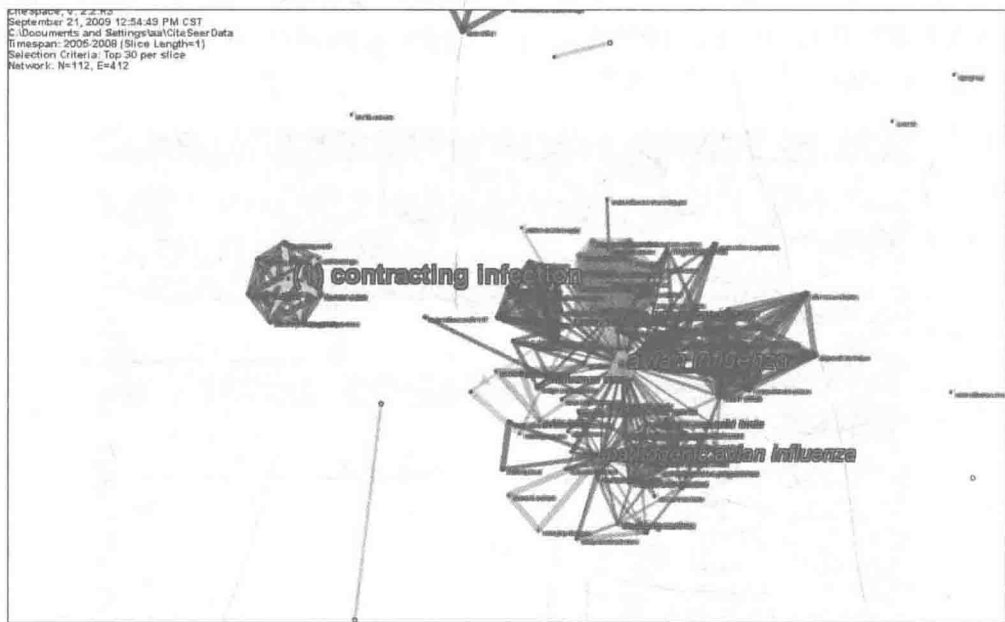


图 11-7 CiteSpace II 禽流感研究前沿术语的共现网络图

(三) 基于文献的知识发现

1. 基于文献的知识发现的定义 对文献信息的分析不仅仅可以汇总、显示已有的研究活动, 甚至还可以通过已发表的文献来发现新的知识, 这个过程被称为基于文献的知识发现 (literature based discovery, LBD)。它是一个将表面上没有任何联系的文献中的具有隐含逻辑关系的知识片段组织起来的信息处理过程。

2. 基于文献的知识发现产生的背景 对文献信息的分析之所以发展到知识发现的深度,根本上是受到了信息爆炸的驱动,同时也是科学发展的必然:一方面,科学研究的深入,造成了客观知识的总量与人类吸收知识的能力之间的差距越来越大。另一方面,科学技术越来越专业化,跨学科的信息传递变得更加困难,就会产生知识的分裂。因此,一个专业领域的信息可能对另一个专业领域是有价值的,跨学科间一定存在着潜在的未被发现的关联。事实上文献间隐性的关联数量可能远多于显性的相互引用的关联数,并且这种隐性关联的发现比信息本身的增长更有意义。

3. 基于文献的知识发现的原理 美国芝加哥大学的信息科学荣誉教授 Don R. Swanson 于 1986 年首次在医学文献研究中发现了这种关系的存在:雷诺病(A)是一种治疗方法和病因都未知的血液循环紊乱疾病,有文献中记载了部分雷诺病患者血液中有某种异常,如血液黏度升高(B)。同时,又在其他文献中发现食用鱼油(C)能纠正这种异常,例如它可以降低血液黏度。因此,Swanson 把这两种知识联系起来得出食用鱼油(C)应该对雷诺病(A)患者有帮助的假设。在这个假说提出大约两年后,有人通过临床实验证实了这一点。

在成功预言了镁缺乏与神经系统疾病(1988年)的关系之后,Swanson 利用自己开发的软件 ARROWSMITH 又发现了很多具有隐藏联系的例子:吡啶美辛与阿尔茨海默病(1996年)、雌激素与阿尔茨海默病(1996年)、游离钙磷脂酶 A2 与精神分裂症(1998年)之间的关系,并发现了可以作为生物武器的潜在病毒(2001年)。他的研究成果引起了人们极大的兴趣。人们首次认识到从文献中可以发现或者挖掘到以前未知的知识。这种经典算法被称作为闭合式知识发现的模式:即从 A 和 C 出发,寻找共同的中间词 B(图 11-8)。

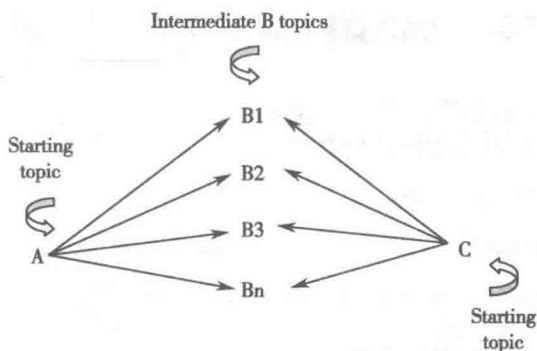


图 11-8 闭合式知识发现模式

近年来,随着研究的深入,人们又提出了开放式知识发现的模式(图 11-9),并开发出可以在网上免费使用的系统。开放式知识发现的过程是,对某个初始研究主题 A,在 MEDLINE 的标题字段检索其相关文献,寻找与 A 在标题中共同出现的中间词 B,通过筛选得到有一定意义

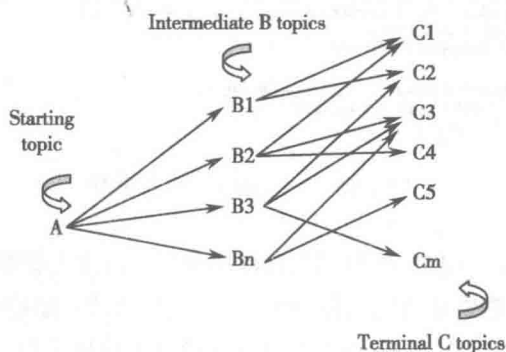


图 11-9 开放式知识发现模式

的 B, 重复上述过程, 得到目标词 C。开放式挖掘初衷是从某个疾病或药物初始, 寻找疾病的潜在病因和治疗方法或者药物的潜在治疗应用。例如输入单个的概念(疾病 A), 找到该概念的第一层相关概念并加以归类, 从中选出药物类的概念, 从第一层相关概念(如药物 B)出发, 找到它们的相关概念 C, 并加以归类(如基因类概念), 然后检验获取的基因和疾病是否有文献报道过。如果没有, 则意味着该基因与疾病有潜在的联系而且并没有文献报道, 这些线索可以提示与疾病、生理学反应或者其他表型相关的新基因、药物或者神经生理过程。

4. 基于文献的知识发现的工具 可以用于基于文献的知识发现的工具主要是用于发现和表示文献或者概念之间潜在联系。

(1) Arrowsmith: 由 D. Swanson 等开发的一个可以在网上免费运行的软件, 通过这个软件可以将 A 和 C 两个主题的文献集中共有的标题词列举出来, 由此来反映两个表面上无联系的事件或者联系微弱的事件(偏头痛与镁)之间的潜在的联系(癫痫)。具体做法是:

首先分别用 A 和 C 两个词进行 MEDLINE 的标题词检索, 将检索到的文献标题分别存储。下一步则是到 ARROWSMITH 的网页上, 把这两个文件的名字输入到相应的框中, 同时上传这两个文件, 随后该系统会产生一个列表(B-LIST), 表中列出在这两个文件所存储的标题中同时出现的重要单词或者词组。这个表中的每一个词都有可能成为那个神秘的 x。用户根据自己的经验把一些没有意义的、含糊的单词去掉。最后, ARROWSMITH 把 B-LIST 中的单词都以链接的方式表现出来, 点击任何一个 x 都会显示出相应的同时包含 A 和 x 的标题, 下面一行会显示出同时含有 x 和 C 的标题。从这些标题中, 用户结合自己的专业知识就可以从中发现主题之间的潜在的联系(图 11-10)。

The screenshot shows the Arrowsmith web interface. At the top, there are tabs for 'Start', 'C-Literature', 'B-list', 'Filter', and 'Literature'. Below these, there is a search bar with 'PubMed' selected and 'gout' entered. A 'Search' button is next to it. Below the search bar, there is a 'History' section showing '1: Search gout : 9621' and a 'Select from History' button. A checkbox labeled 'Use this query for ARROWSMITH?' is checked. Below this, it says 'Items 1 - 20 of 9621' and 'Page 1 of 482'. There is a 'Prev Page' button. The main content area lists five search results, each with a title, authors, journal information, and PMID. The results are: 1. Optimizing Therapy With Allopurinol: Factors Limiting Hypouricemic Efficacy. 2. Effect of sensory education on willingness to taste novel food in children. 3. Comparing the prevalence of rheumatic diseases in China with the rest of the world. 4. EFFICACY OF RASBURICASE THERAPY IN OBSTRUCTIVE NEPHROPATHY FROM RENAL CALCULI: A NOVEL THERAPEUTIC OPTION. 5. Disease-related and all-cause health care costs of elderly patients with gout.

图 11-10 Arrowsmith 交互界面

Swanson 认为 ARROWSMITH 的目标是帮助生物医学者发现新的、有用的联系, 可以把它当做一种扩展 MEDLINE 搜索能力的软件, 但不能替代传统的文献检索, 因为它还必需在传统的文献检索的基础上开展工作。ARROWSMITH 系统目前可以在网上免费使用, 网址为 kiwi.Uchicago.edu。



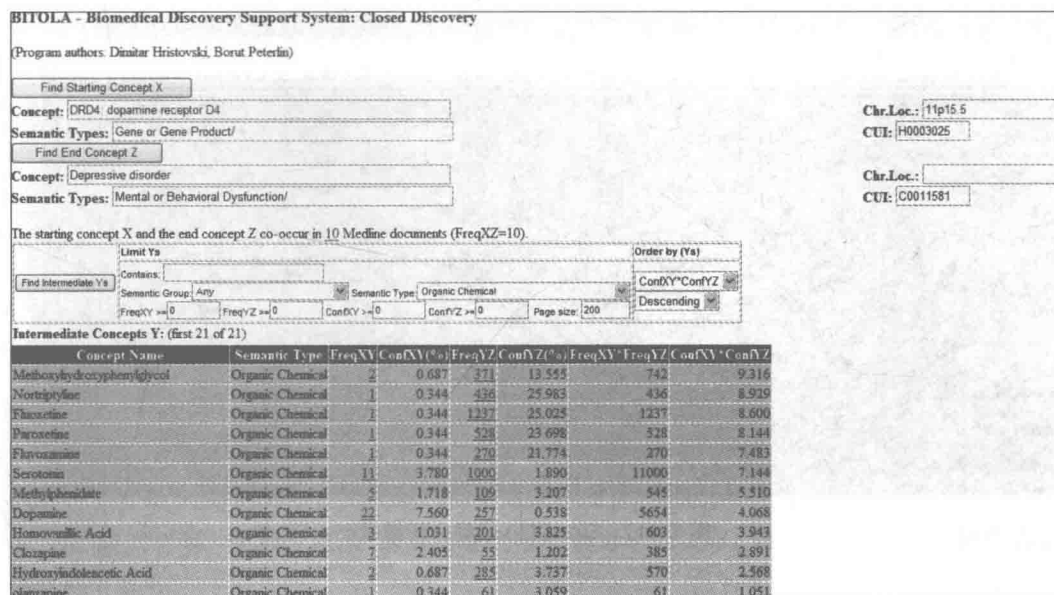


图 11-11 BITOLA 挖掘结果界面

(2) BITOLA: 由斯洛文尼亚的生物统计与医学信息研究所的 Dimitar Hristovski 与 Borut Peterlin 研制开发, 为基于文献的交互式生物医学发现支持系统, 主要用于帮助生物医学研究人员发现生物医学概念之间的潜在联系。网址为 www.mf.uni-lj.si/bitola。

目前系统中采用 MeSH 表中的主题词来表达概念, 同时还包括了来自人类基因组组织 (Human Genome Organisation, HUGO) 的人类基因名称 (<http://www.genenames.org/>), 通过挖掘 Medline 数据库来发现潜在的新关系。该系统采用了闭合式和开放式两种发现模式。例如在闭合式发现中, 用户可以获得一组中间概念 (B) 的列表以及各个概念相应的属性参数, 同时也可以事先限定 B 列表的概念类别。

(3) GenClip: 全称为 Gene Cluster with Literature Profiles, 由南方医科大学肿瘤研究所的黄仲曦等人开发。GenCLiP 工作的原理来自于 Chaussabel D 提出的根据文献轮廓分析基因功能的方法, 即首先检索到与各种基因相关的论文, 统计各种基因与某些重要关键词在同一篇文章中出现的频次, 最终发现基因与这些关键词之间具有生物学意义的关联。

GenCLiP 集成了文献检索、自动抽取关键词、聚类分析、随机模拟、构建基因关系网络图及可视化功能, 帮助用户直观、实时地发掘、理解大量基因数据背后的知识。GenCLiP 有如下两大特点: ①所挖掘出来的疾病机理总是最新的, 因为基因与关键词的关系总是重新挖掘的, 而不是库存的; ②所挖掘出来的信号通路 (网络) 是基于待分析基因和指定关键词的, 而其他大多数软件是针对全部人类基因构建网络, 所得结果不是专门针对待分析的基因。

具体操作是: 先从 PubMed 下载与基因有关的文章 (每个基因最多下载 44 000 篇文章), 并提取出超过设定滤过阈值的关键词 (自动滤除无意义的词汇), 以此为基础对这些基因、关键词建立共词矩阵, 再通过平均距离等级聚类算法进行聚类分析, 寻找基因与特定关键词之间生物学上有显著性意义的关联, 关键词出现达 25% 表示二者强相关, 得到聚类分析点阵图, 如图 11-12。横轴代表用到的关键词, 纵轴代表对应的基因, 图中红色亮点代表该词在对应基因的相关文献中的出现频率。然后, 利用 GenCLiP 的搜索功能, 用户不仅可以得到一个基于特定关键词的与某一基因相关联的基因关系网络图, 还可以查阅到该基因涉及的相关文献。该网络图以单个基因为节点, 与其相关联的基因作为邻近节点, 粉色节点为与特定关键词相关的基因, 蓝色为不相关基因, 节点间连线上的数字代表相邻两种基因相关文献的数目。

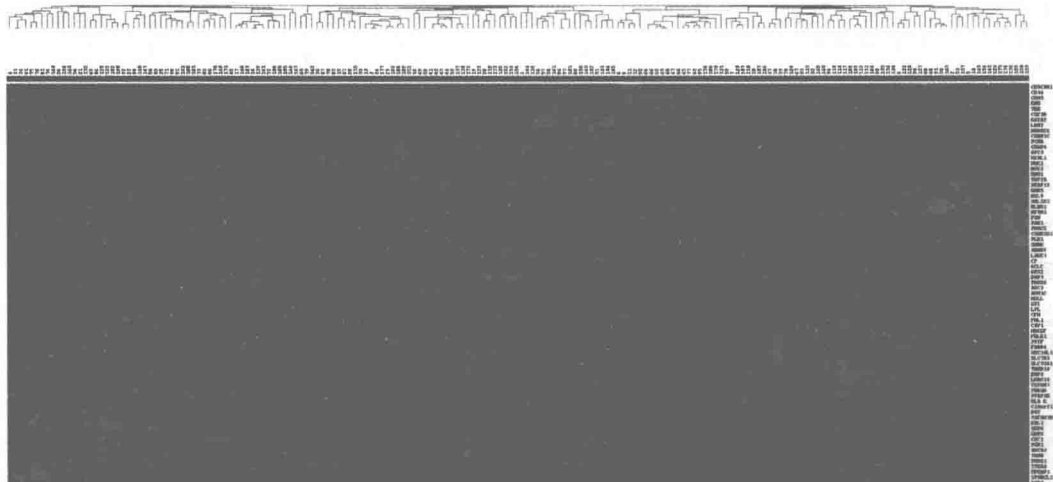


图 11-12 GenCLiP 聚类分析结果

GenCLiP 能简便而有效地用于基因表达谱数据的分析,用来解释复杂的表达数据。该工具通过对大量的基因进行功能聚类使数据变得可以理解,为发现基因间隐含的、有价值的信息提供了机会。目前版本为 2.0,网址为 <http://ci.smu.edu.cn/GenCLiP2.0/analysis.php>。

(崔 雷)

第三节 文献信息的利用

一、文献信息利用的层次与基本原则

文献信息利用,是指文献信息用户对文献信息吸收和运用的活动与过程。文献信息利用是文献信息搜集、整理、评价、分析和研究等环节的延续,同时也是在文献信息搜集、整理、评价、分析和研究等环节的基础上得以完成的。文献信息利用的直接意义主要有两方面:一是实现文献信息存在和运动的社会价值,二是达到文献信息用户进行一系列信息活动的最终目标。

(一) 文献信息利用的层次

由于文献信息的繁杂、用户科学文化水平的不同和社会活动的多样性,使得文献信息利用呈现出层次性。

1. 浅层利用 用户只从表层粗浅地接触文献信息,不求深入地掌握其内容实质,用户可以获得某些表面的、零散的知识 and 消息。例如,为了解国际、国内、本地的新闻和时事而阅读报刊。

2. 中层利用 用户较深入地利用文献信息,力求掌握其内容实质,用户可以获得比较实际的、系统的知识。例如,为解决某些具体问题而到参考工具中寻找答案,为提高医疗教学质量而不断利用新书刊补充教学内容,为提高业务技能利用书刊进行自我学习。

3. 深层利用 用户创造性地利用文献信息,不但深入掌握内容实质而且对其进行分析和研究,用户不但利用已知而且创造新知探索未知。例如,为了课题攻关、技术革新、发明研制而利用文献,为撰写学术论文而利用原始文献。

(二) 文献信息利用中的基本原则

1. 主动利用 能够认识到信息是使人原有知识结构发生变化的那一小部分知识,以及信息的多少意味着消除不确定性的多少,需具有利用文献信息的自觉性和敏感性,而这些又取决于用户利用文献信息的信息意识以及对其工作的熟悉程度。

2. 对口利用 为保证所利用的文献信息有助于社会活动的顺利进展和圆满完成,因而要

密切注意所收集到的文献信息针对性和真实性。一切无效、虚假的文献信息都会带来干扰和危害。

3. 适时利用 文献信息利用相对于用户从事的工作需有一个提前量,但要量度适宜,过早则尚未萌生决策之意,过迟就不免失之为明日黄花。

4. 充分利用 用户所获取的文献中往往包含着许多有用的信息点。只有将其充分地分析、挖掘、研究出来,并最大限度地加以吸收和应用,才能更有效地促进其任务的完成。

5. 全面利用 用户所要利用的文献信息,往往来自多种渠道,涉及多个侧面,如果忽视其广泛性和全面性,就可能因为漏掉某个角度或某个细节的内容造成文献信息利用的失败。

6. 提炼利用 用户所获取的文献信息,在许多情况下是零乱的、片段的,如果对其进行一定的加工、整理,概括,归纳,使其系统化、浓缩化、条理化,在利用时会相对省时省力。

7. 科学利用 用户在利用文献信息时应充分把握其内在的规律,采用先进的、实用的、有效的科学方式方法,这对于提高文献利用的效果和质量,会起事半功倍的作用。

8. 经济利用 用户应遵循“最省力原则”,花费最小的成本、最少的时间、取得最佳的文献信息利用效果。

二、科学研究中的文献信息利用

科研的类型有许多,分类的方法也不尽相同。例如高校科研项目主要是指学校申请获准的纵向科研项目、横向合作科研项目以及校级科研项目。按照项目来自校内外的不同,科研项目可分为“校外科研项目与校内科研项目”。按照项目主管部门与合作单位的不同,校外科研项目又可分为“纵向科研项目与横向科研项目”,纵向科研项目是指列入国家各级科研主管部门科研发展计划的项目。横向科研项目是指接受省内外企事业单位委托,或与企事业单位合作的应用研究和开发研究项目。按照项目立项批准单位级别不同,纵向科研项目可分为“国家级、省部级与厅市级科研项目”;校级科研项目是指校级立项项目。按照项目合作方来自国内外的不同,横向合作科研项目可分为“国际合作项目与国内合作项目”。联合国教科文组织对科技活动中研究与发展的定义与分类为三大类:基础研究、应用研究和试验发展。

获取和利用文献信息是科研工作的重要组成部分,也是科研人员的基本功之一。一般估计科研人员获取文献信息的时间约占整个科研工作的 1/2 左右,而利用文献信息则贯穿科研工作的始终。在科研工作的不同阶段,文献信息利用的特点各不相同。

(一) 科研选题阶段

选题有两种方式,一种是经上级单位指派课题,另一种是科研人员自选课题,无论哪种方式都必须对课题的可行性和新颖性进行论证。深入地利用文献信息,能使科研人员明确了解科研课题的概况,并在此基础上确定这一课题的水平、意义及所在其领域的影响。

(二) 计划制订阶段

科研规划和计划是科研管理的核心,也是科研工作的保证。在制订科研规划和计划时,需要时刻掌握和了解该领域的进展和最新成果,以便确定能否把国内外最新科技成果用在自己的研究项目中。在此基础上,应对该领域的文献信息状况作出客观估计,弄清国内外科技界将会出现的趋势或迹象。

(三) 课题进行阶段

在科研选题阶段和计划制订阶段,科研人员虽然阅读和研究了大量文献信息,但作为研究过程的起点毕竟是初步的,课题必须进一步具体化。在课题进行阶段,必须不断地深入研究前人的文献,在坚实的理论和前人工作的基础上,调整科研计划,更新试验方法,启迪自己的思维,并进一步明确课题中包含的问题,透过表面的现象的问题抓住内层的实质的问题,把模糊的不确切的问题变成清晰的确切的问题。在课题进行阶段,可以通过定题服务,定时定期地

对信息进行分析,以求及时掌握最新成果;若遇到某些难题也可通过专家咨询或研究科技动态,以开阔思路,解决难题。

(四) 课题结束阶段

在此阶段主要是针对成果鉴定和课题总结的要求获取和利用相关的背景材料。首先需要大量的文献信息作论据。并往往由科研管理部门组织科技查新(科技情报查新工作是指查新机构根据查新委托人提供的需要查证其新颖性的科学技术内容,通过检索手段,运用综合分析和对比方法,为科研立项、成果、专利、发明等评价提供科学依据的一种情报咨询服务形式”),召开同行专家、工程技术人员、权威部门的有关人员、管理人员、用户单位的技术人员参加的鉴定会,对科研成果的创造性、科学性、实用性作出切合实际的评价。其次是多数的科研成果都以学术论文、研究报告的文献形式表现出来,使科研成果变成社会的知识。在撰写作品过程中,凡参考借鉴或直接引用的文献内容、实验数据等一般以注释、参考文献等形式指明出处。科研工作利用的文献类型是多种多样的,但不同类型的研究其文献信息利用的侧重点有所不同。例如,基础研究工作多利用理论性较强的一次文献,如期刊、学位论文、科技报告、考察与研究报告、会议论文等,其次是一些相关的标准和专利文献;应用研究工作和开发研究工作主要利用有关新产品、新技术、新工艺等方面具体的文献信息,包括技术期刊、标准、专利、产品样本、图纸、技术报告、实用手册等,而对会议论文和学位论文需求较少。

科研工作利用文献信息的特点是:①科学性强:所利用文献信息要求数据准确,事实正确,理论严密,技术可行;②创造性强:要求利用文献信息中的新思想、新材料、新数据,以激发科研人员创造出新理论、新产品、新结论;③专深性强:利用的文献信息范围较窄,对文献信息内容的研究、理解、吸收的程度非常透彻;④阶段性强:在课题的选定、计划、进行、鉴定等不同阶段,利用文献信息的目的和形式各不相同;⑤多样性强:科研工作所利用的文献信息品种多样,并且既注意从正式交流渠道获取文献信息,又重视从非正式交流渠道获取文献信息。

三、管理活动中文献信息利用

信息是管理的要素,没有信息,管理活动就变成了无本之源了,势必造成混乱和低效。从某种意义上来说,管理就是搜集信息和充分发挥信息的效用,将搜集来的最新信息变为行动。

在管理活动中,每一个阶段都与信息以及文献信息密不可分,决策要以信息为依据,计划要以信息为基础,实施要以信息为保证,而且信息又是组织系统发挥最佳功能的有效纽带,是统一员工思想意志的必要手段,是实现民主管理的重要因素。在管理活动中,如果处理的信息不够,就不能在各项活动中考虑到可能发生的主要因素,从而给管理活动的顺利进行带来很大困难。一般来说,不同层次的管理对信息及文献信息利用的特点有所不同。

基层管理的目标较清楚,工作内容细微详尽,多为重复性的日常业务和事物信息的搜集、存贮、处理、保管、检索、输出等,目的是直接产生一种活动或产品的结果。其信息来源主要有外部信息和内部信息,前者主要指在执行决定过程中来自外界的有关信息,后者主要指由高层管理者决策并经中层管理者制订或解释的限制性信息。

中层管理以实现管理功能和管理技术为对象,以支持和执行决策系统的决定和工作为主,是压缩和分解信息的重要环节。其信息来源主要为内部信息,包括有关限制性信息和下级管理层提供的汇总性信息,经过分析对比后向领导和基层提供管理效能方面的情况。

高层管理的主要任务是根据掌握的内部和外部信息,制定决策方案,提出分析对策和执行报告,权衡对策和方案的利弊得失。由于人们的正确认识往往不是短暂就能完成的,因而所做出的方案往往会出现与客观实际不符合的地方,因而高层管理的决策者要随时获得执行者反馈回来的情况,以便采取相应措施,保证目标的实现。

在管理活动中,最实质的环节是决策,即根据预定目标作出行动的计划安排。在决策的每

个阶段都需要大量利用文献信息。

（一）确定目标阶段

首先要进行大量的调查研究工作,包括全面地搜集相关的文献信息。人们在获得一定数量的信息后,就会对周围环境和所要处理问题有一个比较深刻的理解,才能彻底弄清楚问题的性质、特点、目的、范围和影响,找到决策的目标。其次是在掌握大量信息及文献信息的基础上进行预测,确定解决问题的各种尝试和设想,依据各种过去的和现在的信息,进行分析和研究,判断未来的变化状况和发展趋势。

（二）拟定方案阶段

对获取的各种信息及文献信息进行加工、整理、排队,运用分析、评价、综合、推理、判断等逻辑思维方法加以去粗取精,去伪存真,拟定和设计出各种解决问题的方案计划,并明确其可能性,以供决策者选择。

（三）选择方案阶段

这是决策的最后阶段,关系决策的成败,即对各种方案进行全面的技术论证,综合平衡,对比分析,从中选择出最佳方案。选择的方法一般有三类,即经验判断、数学分析和试验,其中哪一类也离不开信息,不管决策者采用哪一种方法,都有赖于对信息掌握的多少和对信息的判断能力。

在管理活动中,无论需要的是内部信息还是外部信息,有相当部分是以文献的形式出现的,主要涉及的有:①各种行政性文献:如上级文件、管理部门的情况通报、咨询部门的信息参考、基层的计划与总结等,其作用主要在于传递上下左右的信息;②各类报刊:其作用在于通报“短、新、快、实”的新闻消息;③科技文献:在任何一种决策过程中,都必须充分考虑科学技术的第一生产力作用;④经济文献:管理工作始终伴随着对经济因素的分析与研究,如关于世界及区域经济形势、国家经济形势及发展趋势、产品和市场情况的报道;⑤各种政策法规:管理工作应做到不违反国家政策、法律,在决策中必须予以重视,如专利法规、环保法规等。

完整的文献信息分析与利用是一个系列化的智力活动过程,包括前后相随且密切相关的若干个环节,即确定规划与定向前提下,信息搜集、整理与评价;信息分析和提炼;文献信息分析与利用产品的形成;文献信息传递、发表及反馈。在这些环节中,前一个环节是后一个环节的基础,后一个环节是在前一个环节基础上的进一步拓展和深化。在文献信息分析与利用过程中,需要充分调动人的信息器官,广泛摄取相关信息进行思维,是一个不断优化思维的智力活动过程。

本章小结

本章论述了批判性的文献信息搜集、识别、整理、积累、评价和分析中涉及的主要概念和方法。介绍了文献信息分析的基本概念、步骤和方法,着重说明了文献计量学的基本定律,知识发现的方法和工具。了解文献信息的利用层次与基本原则,有利于科研工作者更加清晰地认识到科学研究中文献信息利用的特点。

(周晓政)

思考题

1. 如何确定自己的研究方向?
2. 怎么发现和选择研究课题?
3. 搜集信息的渠道有哪些?

4. 课题文献信息搜集的原则有哪些?
5. 文献信息的整理和积累主要采用哪些方法?
6. 什么是文献分析?
7. 文献信息分析的步骤有哪些?
8. 布拉德福定律文字描述的内容是什么?
9. 基于文献的知识发现的基本原理是什么?
10. 科研工作利用文献信息的特点是什么?



一、MeSH 词表树状结构体系的大类和二级类目(2015)

A. Anatomy

- A01. Body Regions
- A02. Musculoskeletal System
- A03. Digestive System
- A04. Respiratory System
- A05. Urogenital System
- A06. Endocrine System
- A07. Cardiovascular System
- A08. Nervous System
- A09. Sense Organs
- A10. Tissues
- A11. Cells
- A12. Fluids and Secretions
- A13. Animal Structures
- A14. Stomatognathic System
- A15. Hemic and Immune Systems
- A16. Embryonic Structures
- A17. Integumentary System
- A18. Plant Structures
- A19. Fungal Structures
- A20. Bacterial Structures
- A21. Viral Structures

B. Organisms

- B01. Eukaryota
- B02. Archaea
- B03. Bacteria
- B04. Viruses
- B05. Organism Forms

C. Diseases

- C01. Bacterial Infections & Mycoses
- C02. Virus Diseases
- C03. Parasitic Diseases

A. 解剖

- A01. 身体各部位
- A02. 肌肉骨骼系统
- A03. 消化系统
- A04. 呼吸系统
- A05. 泌尿生殖系统
- A06. 内分泌系统
- A07. 心血管系统
- A08. 神经系统
- A09. 感觉器官
- A10. 组织
- A11. 细胞
- A12. 体液和分泌物
- A13. 动物结构
- A14. 口颌系统
- A15. 血液与免疫系统
- A16. 胚胎结构
- A17. 皮肤系统
- A18. 植物结构
- A19. 真菌结构
- A20. 细菌结构
- A21. 病毒结构

B. 有机体

- B01. 真核生物
- B02. 古细菌
- B03. 细菌
- B04. 病毒
- B05. 生物形态

C. 疾病

- C01. 细菌感染与真菌病
- C02. 病毒疾病
- C03. 寄生虫病

- C04. Neoplasms
- C05. Musculoskeletal Diseases
- C06. Digestive System Diseases
- C07. Stomatognathic Diseases
- C08. Respiratory Tract Diseases
- C09. Otorhinolaryngologic Diseases
- C10. Nervous System Diseases
- C11. Eye Diseases
- C12. Male Urogenital Diseases
- C13. Female Urogenital Diseases and Pregnancy Complications
- C14. Cardiovascular Diseases
- C15. Hemic and Lymphatic Diseases
- C16. Congenital, Hereditary, and Neonatal Diseases and Abnormalities
- C17. Skin and Connective Tissue Diseases
- C18. Nutritional and Metabolic Diseases
- C19. Endocrine System Diseases
- C20. Immune System Diseases
- C21. Disorders of Environmental origin
- C22. Animal Diseases
- C23. Pathological Conditions, Signs and Symptoms
- C24. Occupational Diseases
- C25. Chemically-Induced Disorders
- C26. Wounds and Injuries

D. Chemicals and Drugs

- D01. Inorganic Chemicals
- D02. Organic Chemicals
- D03. Heterocyclic Compounds
- D04. Polycyclic Compounds
- D05. Macromolecular Substances
- D06. Hormones, Hormone Substitutes, and Hormone Antagonists
- D08. Enzymes and Coenzymes
- D09. Carbohydrates
- D10. Lipids
- D12. Amino Acids, Peptides and Proteins
- D13. Nucleic Acids, Nucleotides and Nucleosides
- D20. Complex Mixtures
- D23. Biological Factors
- D25. Biomedical and Dental Materials
- D26. Pharmaceutical Preparations
- D27. Chemical Actions and uses

- C04. 肿瘤
- C05. 肌肉骨骼的疾病
- C06. 消化系统疾病
- C07. 口颌疾病
- C08. 呼吸道疾病
- C09. 耳鼻喉疾病
- C10. 神经系统疾病
- C11. 眼疾病
- C12. 男性泌尿生殖系统疾病
- C13. 女性泌尿生殖系统疾病与妊娠合并症
- C14. 心血管系统疾病
- C15. 血液与淋巴疾病
- C16. 先天性、遗传性、新生儿疾病与畸形
- C17. 皮肤与结缔组织疾病
- C18. 营养与代谢疾病
- C19. 内分泌系统疾病
- C20. 免疫系统疾病
- C21. 环境因素诱发疾病
- C22. 动物疾病
- C23. 病理状态、体征与症状
- C24. 职业病
- C25. 化学物质诱导疾病
- C26. 创伤与损伤

D. 化学品与药物

- D01. 无机化合物
- D02. 有机化合物
- D03. 杂环化合物
- D04. 多环化合物
- D05. 大分子物质
- D06. 激素、激素代用品与激素拮抗剂
- D08. 酶与辅酶
- D09. 糖类
- D10. 脂类
- D12. 氨基酸、肽与蛋白质
- D13. 核酸、核苷酸与核苷
- D20. 复合物
- D23. 生物因子
- D25. 生物医学与牙科材料
- D26. 药物制剂
- D27. 化学活性与应用

E. Analytical, Diagnostic and Therapeutic Techniques and Equipment

- E01. Diagnosis
- E02. Therapeutics
- E03. Anesthesia and Analgesia
- E04. Surgical, procedures, Operative
- E05. Investigative Techniques
- E06. Dentistry
- E07. Equipment and Supplies

F. Psychiatry and Psychology

- F01. Behavior and Behavior Mechanisms
- F02. Psychological Phenomena and Processes
- F03. Mental Disorders
- F04. Behavioral Disciplines and Activities

G. Phenomena and Processes

- G01. Physical Phenomena
- G02. Chemical Phenomena
- G03. Metabolic Phenomena
- G04. Cell Physiological Phenomena
- G05. Genetic Phenomena
- G06. Microbiological Phenomena
- G07. Physiological Phenomena
- G08. Reproductive and Urinary Physiology Phenomena
- G09. Circulatory and Respiratory Physiology Phenomena
- G10. Digestive System and Oral Physiology Phenomena
- G11. Musculoskeletal and Neural Physiology Phenomena
- G12. Immune System Phenomena
- G13. Integumentary System Physiological Phenomena
- G14. Ocular Physiological Phenomena
- G15. Plant Physiological Phenomena
- G16. Biological Phenomena
- G17. Mathematical Concepts

H. Disciplines and Occupations

- H01. Natural Science Disciplines
- H02. Health Occupations

I. Anthropology, Education, Sociology and Socials Phenomena

- I01. Social Sciences
- I02. Education
- I03. Human Activities

J. Technology, Industry, Agriculture

- J01. Technology, Industry and Agriculture
- J02. Food and Beverages

E. 分析、诊断、治疗技术与设备

- E01. 诊断
- E02. 治疗
- E03. 麻醉与镇痛
- E04. 外科操作, 手术
- E05. 包埋技术
- E06. 牙科
- E07. 设备与供应

F. 精神病学与心理学

- F01. 行为与行为机制
- F02. 心理现象与过程
- F03. 精神疾病
- F04. 行为科学与活动

G. 现象与操作

- G01. 物理现象
- G02. 化学现象
- G03. 代谢现象
- G04. 细胞生理现象
- G05. 遗传现象
- G06. 微生物现象
- G07. 生理现象
- G08. 生殖和泌尿生理现象
- G09. 循环和呼吸生理现象
- G10. 消化系统与口腔生理现象
- G11. 肌肉骨骼与神经生理现象
- G12. 免疫系统现象
- G13. 表皮系统生理现象
- G14. 眼生理现象
- G15. 植物生理现象
- G16. 生物学现象
- G17. 数学概念

H. 学科与职业

- H01. 自然科学
- H02. 医疗卫生职业

I. 人类学、教育、社会学与社会现象

- I01. 社会科学
- I02. 教育
- I03. 人类活动

J. 技术、工业、农业

- J01. 技术、工业与农业
- J02. 食物与饮料

K. Humanities

K01. Humanities

L. Information Science

L01. Information Science

M. Named Groups

M01. Persons

N. Health Care

N01. Population Characteristics

N02. Health Care Facilities, Manpower and Services

N03. Health Care Economics and Organizations

N04. Health Services Administration

N05. Health Care Quality, Access and Evaluation

N06. Environment and Public Health

V. Publication Characteristics

V01. Publication Components

V02. Publication Formats

V03. Study Characteristics

V04. Support of Research

Z. Geographicals

Z01. Geographic Locations

K. 人文科学

K01. 人文科学

L. 情报科学

L01. 情报科学

M. 命名群组

M01. 人群

N. 卫生保健

N01. 人口特征

N02. 卫生保健设施、人力与服务

N03. 卫生保健经济与组织

N04. 卫生服务管理

N05. 卫生保健质量, 实施与评估

N06. 环境与公共卫生

V. 出版特征

V01. 出版组成

V02. 出版格式

V03. 研究类型

V04. 研究支持

Z. 地理

Z01. 地理位置

二、MeSH 词表限定词使用一览表(2015)

下表按 83 个限定词的英文字母排列, 并非每个限定词都能与所有主题词组配, 每个限定词括号内的主题词的类目范畴是每个限定词进行组配的一般指南。限定词的详细记录及每个主题词能组配的具体限定词可利用 MeSH Browser 查询获取。

1. Abnormalities 畸形 (A1-5, A7-10, A13, A14, A16, A17) AB, abnorm

与器官主题词组配, 表明因先天性缺陷而致器官的形态学改变。亦用于动物的畸形。

2. Administration & Dosage 投药和剂量 (D1-6, D8-10, D12, D13, D20, D23, D25-27) AD, admin

与药物主题词组配, 表明其剂型、给药途径、次数、用药时间、剂量以及这些因素的作用。

3. Adverse Effects 副作用 (A18, D1-6, D8-10, D12, D13, D20, D23, D25-27, E2-4, E6, E7, J2) AE, adv eff

与药物、化学物质、生物制品、物理制剂或各种制品组配, 表明其在以诊断、治疗、预防或麻醉为目的, 正常用量或可接受的剂量情况下所出现的不良反应; 也与各种诊断、治疗、预防、麻醉、手术或其他技术操作组配, 表明因操作而引起的副作用或并发症。不包括禁忌证, 禁忌证副主题词是 Contraindications。

4. Agonists 激动剂 (D1-4, D6, D9, D10, D12, D13, D23) AG, agon

与化学物质、药物、内源性物质主题词组配, 表明这些物质对受体具有亲和力及内在作用。

5. Analogs & Derivatives 类似物和衍生物 (D3) AA, analogs

与药物及化学物质组配, 表明是具有相同或相似电子结构, 但其原子或分子不同(增加或取代)的物质。在 MeSH 表中无此专指的化学物质主题词或合适的化学结构族主题词时使用。

6. Analysis 分析 (D1-6, D8-10, D12, D13, D20, D23, D25-27) AN, anal

用于某种物质的成分或其代谢产物的鉴定或定量测定, 包括对空气、水或其它环境载体的化学分析, 但不包括对组织、肿瘤、体液、有机体和植物的化学分析。对后者使用副主题词“化学”。本概念可用于

分析的方法学和结果。血液、脑脊髓液和尿中的物质分析,分别用副主题词“血液(Blood)”、“脑脊髓液(Cerebrospinal Fluid)”和“尿(Urine)”。

7. Anatomy & Histology 解剖学和组织学(A1-5, A7-10, A13, A14, A16-18, B1)AH, anat

与器官、部位及组织组配,表明其正常解剖学和组织学。也与动、植物组配,表明其正常的解剖学及结构。

8. Antagonists & Inhibitors 拮抗剂和抑制剂(D2-6, D8-10, D12, D13, D23)AI, antag

与药物、化学物质、内源性物质主题词组配,表明在生物效应上与其有相反作用机制的物质或制剂。

9. Biosynthesis 生物合成(D6, D8, D9, D12, D13, D23)BI, biosyn

与化学物质组配,表明其在有机体内、活细胞内或亚细胞成分中的形成。

10. Blood 血液(C1-26, D1-4, D6, D8-10, D12, D13, D20, D23, D27, F3)BL, blood

用于表明血液中各种物质的存在或分析。也用于疾病状态时的血液检查和血液变化。但不包括血清诊断和血清学。前者用副主题词“诊断 diagnosis”,后者用“免疫学 immunology”。

11. Blood Supply 血液供给(A1-5, A8-10, A13, A14, A16, C4)BS, blood supply

与器官,身体部位组配,在需与血管主题词组配时,如无专指的血管主题词,可与某器官、部位的动脉、毛细血管及静脉系统组配,表明通过器官内的血流。

12. Cerebrospinal Fluid 脑脊髓液(C1-26, D1-4, D6, D8-10, D12, D13, D20, D23, D27, F3)CF, csf

表明脑脊髓液中物质的存在和分析,也用于疾病状态时,脑脊髓液的检查和变化。

13. Chemical Synthesis 化学合成(D2-6, D8-10, D12, D13, D20, D25-27)CS, chem syn

与化学物质和药物主题词组配,表明体外分子的化学制备。有机体内、活细胞内或亚细胞内化学物质的形成,用副主题词“生物合成(Biosynthesis)”。

14. Chemically Induced 化学诱导(C1-20, C22, C23, C26, F3)CI, chem ind

由内源性或外源性物质引起的生物现象、疾病、综合征、先天性畸形或症状。

15. Chemistry 化学(A2-21, B1-5, C4, D1-6, D8-10, D12, D13, D20, D23, D25-27)CH, chem

与化学、生物或非生物物质组配,表明其组成结构、特点和性质。也用于器官、组织、肿瘤、体液、有机体和植物的化学成分和物质含量。但不包括物质的化学分析和测定、合成、分离和提纯,对后几种情况,分别用副主题词“分析(Analysis)”、“化学合成(Chemical Synthesis)”、“分离和提纯(Isolation & Purification)”。

16. Classification 分类(A11, A15, A18-21, B1-5, C1-26, D1-6, D8-10, D12, D13, D20, D23, D25-27, E1-7, F3, H1, H2, I2, I3, J1, J2, M1, N2-4)CL, class

用于分类学的或体系的或等级的分类系统。

17. Complications 并发症(C1-26, F3)CO, compl

与疾病组配,表明两种病同时存在或相继存在的状况,即同时存在的疾病、并发症或后遗症。

18. Congenital 先天性(C1-12, C14, C15, C17, C19-26)CN, congen

与疾病组配,表明出生时或出生前即存在的疾病。但不包括形态畸形和分娩时的损伤,后两者分别用副主题词“畸形(Abnormalities)”和“损伤(Injuries)”。

19. Contraindications 禁忌证(D1-6, D8-10, D12, D13, D20, D23, D25-27, E2-4, E6, E7)CT, contra

与药物、化学物质、生物和物理因子组配,表明在疾病或生理状态下,使用这些物质可能不合适、不需要或不可取,也适应于诊断、治疗、预防、麻醉、外科或者其他操作的禁忌证。

20. Cytology 细胞学(A2-10, A12-19, B1-3, B5)CY, cytol

用于单细胞或多细胞有机体的正常细胞形态学。

21. Deficiency 缺乏(D6, D8, D12)DF, defic

与内源性和外源性物质组配,表明其缺乏或低于有机体或生物系统的正常需要量。

22. Diagnosis 诊断(C1-26, F3)DI, diag

与疾病组配,表明诊断的各个方面,包括检查、鉴别诊断及预后。但不包括普查、放射照相诊断、放射

性核素成像、超声诊断, 对后几种情况, 分别用副主题词“预防和控制(Prevention & Control)”、“放射照相术(Radiography)”、“放射性核素成像(Radionuclide Imaging)”、“超声诊断(Ultrasonography)”。

23. Diagnostc Use 诊断应用(D1-6, D8-10, D12, D13, D20, D23, D25-27)DU, diag use

与化学物质、药物和物理因子组配, 表明其对器官临床功能的研究, 或对人类或动物疾病的诊断。

24. Diet Therapy 饮食疗法(C1-26, F3)DH, diet ther

与疾病组配, 表明对疾病所作的饮食和营养安排。但维生素或矿物质的补充, 则用副主题词“药物治法(Drug Therapy)”。

25. Drug effects 药物作用(A2-21, B1-5, D8, D12, D23, G2-15)DE, drug eff

与器官、部位、组织或有机体以及生理和心理过程组配, 表明药品和化学物质对其产生的作用。

26. Durg Therapy 药物疗法(C1-26, F3)DT, drug ther

与疾病组配, 表明通过投给药品、化学物质和抗生素治疗疾病。但不包括免疫治疗和用生物制品治疗, 对后者用副主题词“治疗(Therapy)”。对于饮食疗法和放射疗法, 分别用专指的副主题词。

27. Economics 经济学(C1-26, D1-6, D8-10, D12, D13, D20, D23, D25-27, E1-7, F3, H1, H2, I2, I3, J1, J2, N2-4)EC, econ

用于任何主题的经济方面。也用于金融管理的各个方面, 包括资金的筹集和提供。

28. Education 教育(E4, H1, H2, M1)ED, educ

与学科、技术和人群主题词组配, 表明各个领域和学科以及各类人群的教育和培训。

29. Embryology 胚胎学(A1-5, A7-10, A13, A14, A16-18, B1, C1-26)EM, embryol

与器官、部位和动物主题词组配, 表明其在胚胎期或胎儿期的发育。也与疾病主题词组配, 表明因胚胎因素而引起的出生后疾病。

30. Enzymology 酶学(A2-21, B1-5, C1-26, F3)EN, enzymol

与有机体(脊椎动物除外)、器官和组织、疾病主题词组配, 表明有机体、器官组织的酶以及疾病过程中的酶。但不包括诊断性酶试验, 后者用副主题词“诊断(Diagnosis)”。

31. Epidemiology 流行病学(C1-26, F3, Z1)EP, epidemiol

与人类或兽医疾病组配, 表明疾病的分布、致病因素和特定人群的疾病特征, 包括发病率和患病率和发病周期、地方病和流行病暴发流行, 包括对地理区域和特殊人群发病率的调查和估计。也与地理主题词组配, 表明疾病流行病学方面的地理位置。但不包括死亡率, 死亡率用副主题词“死亡率(Mortality)”。

32. Ethics 伦理学(E1-7, G9, H1, H2, I2, I3, N2-4)ES, ethics

与讨论、分析有关人类和社会价值的技术与活动的主题词组配。

33. Ethnology 人种学(C1-21, C23-26, F3, Z1)EH, ethnol

与疾病和有关主题词组配, 表明疾病的人种的、文化的、人类学的或种族的等方面。也与地理主题词组配, 表明人群的起源地。

34. Etiology 病因学(C1-26, F3)ET, etiol

与疾病组配, 表明疾病的致病原因(包括微生物、环境因素、社会因素和个人习惯)及发病机制。

35. Genetics 遗传学(A18-21, B1-5, C1-26, D6, D8, D12, D13, D23, F3, G2-15)GE, genet

与有机体主题词组配, 表明其遗传和遗传机制, 正常和病理状态下的遗传学基础。也用于内源性化学物质的遗传学方面的研究; 包括对遗传物质的生物化学和分子影响。

36. Growth & Development 生长和发育(A1-5, A7-10, A13, A14, A16-19, B1-5)GD, growth

与微生物、植物和出生后动物主题词组配, 表明其生长和发育情况。也与器官和解剖部位主题词组配, 表明其出生后的生长和发育情况。

37. History 历史(C1-26, D1-6, D8-10, D12, D13, D20, D23, D25-27, E1-7, F3, F4, H1, H2, I1-3, J1, J2, M1, N1-4)HI, hist

用于任何主题词组配, 表明其历史情况, 包括简要的历史注释, 但不包括病史。

38. Immunology 免疫学(A2-21, B1-5, C1-26, D1-6, D8-10, D12, D13, D20, D23, D27, F3, G3-15) IM, immunol

与组织、器官、微生物、真菌、病毒和动物主题词组配,包括疾病的免疫学方面。但不包括用于诊断、预防和治疗为目的的免疫学操作,对后者分别用副主题词“诊断(Diagnosis)”、“预防和控制(Prevention & Control)”、“治疗(Therapy)”。与化学物质主题词组配,指作为抗原和半抗原的化学物质。

39. Injuries 损伤(A1-5, A7, A9, A10, A13, A14, A16, A17) IN, inj

与解剖学、动物和运动主题词组配,表明其所受的创伤和损坏。但不包括细胞损坏,对后者用副主题词“病理学(Pathology)”。

40. Innervation 神经支配(A1-5, A7, A9, A10, A13, A14, A16, A17) IR, innerv

与器官、部位或组织组配,表明其神经分布。

41. Instrumentation 仪器设备(E1-5, H1, H2) IS, instrum

与诊断或治疗操作、分析技术及专业或学科主题词组配,表明器械、仪器或设备的研制和改进。

42. Isolation & Purification 分离和提纯(A21, B2-4, D1-6, D8-10, D12, D13, D20, D23, D25-27) IP, isol

与细菌、病毒、真菌、原生动物和蠕虫主题词组配,表明对其纯株的获取或通过 DNA 分析、免疫学和其他方法(包括培养技术)以显示上述有机体的存在或对其进行鉴定。也与生物物质和化学物质主题词组配,表明对其成分的分离和提纯。

43. Legislation & Jurisprudence 立法和法学(E4, H1, H2, I2, I3, M1, N2-4) LJ, legis

用于法律、法令、条令或政府法规以及法律争议和法庭判决。

44. Manpower 人力(H1, H2) MA, man

与学科和规划项目主题词组配,表明其对人员的需求、提供、分配、招聘和使用。

45. Metaboism 代谢(A2-21, B1-5, C1-26, D1-6, D8-10, D12, D13, D20, D23, D25-27, F3) ME, metab

与器官、细胞和亚细胞成分、有机体和疾病主题词组配,表明其生化改变及代谢情况。也与药品和化学物质组配,表明其合成代谢过程(从小分子到大分子的转换)和分解代谢变化(从复杂分子分解为简单分子)。对于生物合成、酶学、药代动力学和分泌,则另用相应的副主题词。

46. Methods 方法(E1-5, H1, H2) MT, methods

与技术、操作和规划项目主题词组配,表明其方法。

47. Microbiology 微生物学(A1-18, A20, B1, C1-26, E7, F3, J2) MI, microbiol

与器官、动物、高等植物和疾病主题词组配,表明对其进行微生物学研究。对寄生虫和病毒的研究,则用相应的副主题词。

48. Mortality 死亡率(C1-26, E2-4, F3) MO, mortal

与人类疾病和兽医疾病组配,表明其死亡率的统计。由于特殊病例引起的死亡用“致命效果(Fatal outcome)”。

49. Nursing 护理(C1-26, E2-4, F3) NU, nurs

与疾病组配,表明对疾病的护理和护理技术,包括诊断、治疗和预防操作中的护理作用。

50. Organization & Administration 组织和管理(H1, H2, I2, N2, N4) OG, organ

与机构或卫生保健组织主题词组配,表明行政机构及其管理。

51. Parasitology 寄生虫学(A1-18, A20, B1, C1-26, E7, F3, J2) PS, parasitol

与动物、高等植物、器官和疾病主题词组配,表明其寄生虫因素。在疾病诊断过程中,寄生虫因素不明确时勿用该副主题词。

52. Pathogenicity 致病力(B2-4) PY, pathogen

与微生物、病毒和寄生虫主题词组配,表明其对人或动物致病能力的研究。

53. Pathology 病理学(A1-11, A13-17, A20, C1-26, F3) PA, pathol

与器官、组织及疾病主题词组配,表明疾病状态时,器官、组织及细胞的结构。

54. Pharmacokinetics 药代动力学 (D1-6, D8-10, D12, D13, D20, D25-27) PK, pharmacokin
与药品和外源性化学物质组配, 表明以其剂量的效用和代谢过程的扩展和速率研究其吸收、生物转化、分布、释放、转运、摄取和排泄的机制和动力学。
55. Pharmacology 药理学 (D1-6, D8-10, D12, D13, D20, D23, D25-27) PD, pharmacol
与药品和外源性投给的化学物质组配, 表明其对活的组织和有机体的作用, 包括对生理及生化过程的催化和抑制以及其他药理作用机制。
56. Physiology 生理学 (A1-21, B1-5, D6, D8, D12, D13, D23, G2-15) PH, physiol
与器官、组织和单细胞及多细胞有机体组配, 表明其正常功能; 也与生化物质、内源性物质组配, 表明其生理作用。
57. Physiopathology 病理生理学 (A1-5, A7-10, A13, A14, A16, A17, C1-26, F3) PP, physiopathol
与器官和疾病组配, 表明疾病状态下的功能异常。
58. Poisoning 中毒 (A18, D1-6, D8-10, D12, D13, D20, D23, D25-27, J2) PO, pois
与药品、化学物质和工业原料等主题词组配, 表明因上述物质引起的人或动物急、慢性中毒, 包括因意外、职业、自杀、误用以及环境污染等所致的中毒。
59. Prevention & Control 预防和控制 (C1-26, F3) PC, prev
与疾病组配, 表明增加人和动物的抗病能力(如免疫接种), 对传播媒介的控制, 对环境有害因素和致病的社会因素的预防和控制, 包括对个体的预防措施。
60. Psychology 心理学 (C1-26, E2-4, E6, F3, I3, M1) PX, psychol
与非精神性疾病、技术及人群名称主题词组配, 表明其心理学的、精神的、身心的、心理社会学的、行为的和感情的等方面; 也与精神性疾病组配, 表明其心理方面; 也与动物主题词组配, 表明动物行为和心理学。
61. Radiation Effects 辐射作用 (A1-21, B1-5, D1-6, D8-10, D12, D13, D20, D23, D25-27, G2-15, J2) RE, rad eff
与活的有机体、器官和组织及其组成部分、生理过程等主题词组配, 表明电离和非电离辐射对其产生的作用; 也与药品和化学物质组配, 表明辐射对其产生的效应。
62. Radiography 放射照相术 (A1-17, A20, C1-26, F3) RA, radiogr
与器官、部位和疾病主题词组配, 表明对其进行 X 线检查。不包括放射性核素成像, 对后者用副主题词“放射性核素成像 Radionuclide Imaging”。
63. Radionuclide Imaging 放射性核素成像 (A1-17, A20, C1-26, F3) RI, radionuclide
与解剖学和疾病主题词组配, 表明对任何解剖结构的放射性成像以及对疾病的诊断。
64. Radiotherapy 放射疗法 (C1-26) RT, radiother
与疾病主题词组配, 表明电离和非电离辐射的治疗作用, 包括放射性同位素疗法。
65. Rehabilitation 康复 (C1-21, C23-26, E4, F3) RH, rehabil
与疾病和外科手术操作主题词组配, 表明(病后或术后)个体功能的恢复。
66. Secondary 继发性 (C4) SC, second
与肿瘤主题词组配, 表明肿瘤转移的继发部位。
67. Secretion 分泌 (A3-16, A20, C4, D6, D8, D12, D13) SE, secret
与器官、组织、腺体、肿瘤和内源性物质主题词组配, 表明由于器官、组织或腺体的完整细胞活动而产生的内源性物质, 经细胞膜排出进入细胞间隙或管内。
68. Standards 标准 (D1-6, D8-10, D12, D13, D20, D25-27, E1-7, F4, H1, H2, I2, J1, J2, N2-4) ST, stand
与设施、人员和规划项目主题词组配, 表明对其合适的或可行的标准的制订、测试或应用; 也与化学物质和药品组配, 表明其鉴定标准、质量标准和效率标准, 包括工业和职业中的卫生和安全标准。
69. Statistics & Numerical Data 统计和数值数据 (E1-7, F4, H1, H2, I2, I3, M1, N2-4) SN, statist
与非疾病主题词组配, 表明对数值的表达, 即对特定的数值集合或数值组进行描述。不包括人力分配

和物质设备等, 对后两种情况, 分别用副主题词“人力 (Manpower)”和“供应和分配 (Supply & Distribution)”。

70. Supply & Distribution 供应和分配 (D1-6, D8-10, D12, D13, D20, D25-27, E7, J2) SD, supply
与物质、设备、卫生保健服务、人员和设施主题词组配, 表明可能获得上述项目的数量和分布情况。但不包括企业、事业单位中的食品和水的供应。

71. Surgery 外科手术 (A1-5, A7-10, A13, A14, A16, C1-26, F3) SU, surg
与器官、部位、组织和疾病主题词组配, 表明以手术治疗疾病, 包括用激光切除组织。但不包括移植术, 对后者用副主题词“移植 (Transplantation)”。

72. Therapeutic Use 治疗应用 (D1-6, D8-10, D12, D13, D20, D23, D25-27) TU, ther use
与药品、生物制品和物理因子主题词组配, 表明其在疾病预防和治疗中的应用; 也包括兽医用药。

73. Therapy 治疗 (C1-26, F3) TH, ther
与疾病主题词组配, 用于除药物治疗、饮食疗法、放射疗法和外科手术以外的治疗手段。包括综合治疗。

74. Toxicity 毒性 (A18, D1-6, D8-10, D12, D13, D20, D23, D25-27, J2) TO, tox
与药品及化学物质主题词组配, 表明其对人和动物有害作用的实验性研究, 如安全剂量的测定, 以及按不同剂量给药产生的不良反应; 也用于接触环境污染物的实验研究。

75. Transmission 传播 (C1-3, C22) TM, transm
与疾病主题词组配, 表明对疾病传播方式的研究。

76. Transplantation 移植 (A2, A3, A5-11, A13-17, A20) TR, transpl
与器官、组织或细胞主题词组配, 表明器官、组织或细胞在同一个体中由一个部位移植到另一个部位, 或在同种或异种不同个体间的移植。

77. Trends 发展趋势 (E1-7, H1, H2, I2, I3, N2-4) TD, trends
用于表明事物随时间的推移而发生质变和量变的方式, 包括过去、现在和未来的情况。但不包括对具体病人的疾病过程的讨论。

78. Ultrasonography 超声检查 (A1-17, A20, C1-26, F3) US, ultrasonogr
与器官、部位主题词组配, 表明对其进行超声成像; 也与疾病主题词组配, 表明对疾病进行超声诊断, 但不包括超声治疗。

79. Ultrastructure 超微结构 (A2-11, A13-21, B1-5, C4, D8, D12) UL, ultrastruct
与组织和细胞 (包括肿瘤) 和微生物主题词组配, 表明其通常用光显微镜观察不到的细微解剖结构。

80. Urine 尿 (C1-26, D1-4, D6, D8-10, D12, D13, D20, D23, D27, F3) UR, urine
表明尿液中物质的存在与分析; 也表明疾病状态时, 尿液中物质的变化及尿液检查。

81. Utilization 利用 (E1-4, E6, E7, N2, N4) UT, util
与设备、设施、规划项目、服务和卫生人员主题词组配, 讨论其利用情况 (通常用数据); 也包括讨论利用过度 and 利用不足。

82. Veterinary 兽医学 (C1-21, C23-26, E1-4, E6, E7) VE, vet
与疾病主题词组配, 表明动物自然发生的疾病; 也与技术操作主题词组配, 表明兽医学中使用的诊断、预防或治疗操作。

83. Virology 病毒学 (A1-20, B1-3, B5, C1-26, E7, F3, J2) VI, virol
与器官、动物、高等植物以及疾病的主题词组配, 表明病毒学研究。而对细菌、立克次体以及真菌等微生物的研究, 则用副主题词“微生物 (Microbiology)”, 寄生虫方面的研究用“寄生虫学 (Parasitology)”。

参考文献及推荐网站

1. 方平. 医学文献信息检索. 北京: 人民卫生出版社, 2005.
2. 罗爱静. 医学文献信息检索. 第2版. 北京: 人民卫生出版社, 2010.
3. 杨克虎. 生物医学信息检索与利用. 北京: 人民卫生出版社, 2009.
4. 罗爱静, 胡德华. 医学科技信息检索. 长沙: 中南大学出版社, 2008.
5. 徐一新. 医学信息检索. 北京: 高等教育出版社, 2004.
6. 沙勇忠. 信息伦理学. 北京: 北京图书馆出版社, 2004.
7. 杨玉圣, 张保生. 学术规范导论. 北京: 高等教育出版社, 2004.
8. 于双成. 科技信息检索与利用. 北京: 清华大学出版社, 2011.
9. 王庭槐. 医学电子资源获取与利用. 北京: 高等教育出版社, 2013.
10. 肖珑. 数字信息资源的检索与利用. 第2版. 北京: 北京大学出版社, 2013.
11. 李晓玲, 符礼平. 医学信息检索与利用. 第5版. 上海: 复旦大学出版社, 2014.
12. 乔纳森·佩夫斯纳. 生物信息学与功能基因组学. 北京: 化学工业出版社, 2006.
13. 李霞. 生物信息学. 北京: 人民卫生出版社, 2010.
14. 代涛. 医学信息搜集与利用. 第2版. 北京: 人民卫生出版社, 2014.
15. 王庭槐. MOOC-席卷全球教育的大规模开放在线课程. 北京: 人民卫生出版社, 2014/8/25.
16. 田运. 信息与思维. 福州: 福建教育出版社, 1990.
17. 乔好勤等. 文献信息检索与利用. 第2版. 武汉: 华中科技大学出版社, 2013.
18. 王延飞等. 信息分析与决策. 北京: 北京大学出版社, 2010.
19. 张怀涛等. 文献信息利用通论. 北京: 原子能出版社, 1996.
20. (英)斯特拉·科特雷尔. 批判性思维训练手册. 北京: 北京大学出版社, 2012.
21. 朱庆华. 信息分析: 基础、方法及应用. 北京: 科学出版社, 2004.
22. 邱均平. 信息计量学. 武汉: 武汉大学出版社, 2007.
23. 赖茂生等. 计算机情报检索, 第2版. 北京: 北京大学出版社, 2006, 10.
24. 皮介郑. 信息素质理论与教育研究. 北京: 中国科学院研究生院文献情报中心, 2003.
25. 马海群. 论信息素质教育. 中国图书馆学报, 1997, (2).
26. 李晓娟. 美国高校中信息素质教育项目的实施及经验分析[J]. 中国图书馆学报, 2003 (2).
27. 孙建军, 成颖, 张玲玲. 信息素质教育进展. 情报学报, 2003, 22 (2).
28. 余恩琳. 医学生信息素质教育研究. 医学教育探索, 2005 (6).
29. 陈晨, 高洁. 美国犹他州大学图书馆联盟网络导航课程剖析. 图书馆杂志. 2012 (10): 77-79+46.
30. 于双成, 张晓一, 徐丽梅. 临床医生医学素养结构的哲学解析. 医学与哲学, 2008, 29 (3): 50-51.
31. 于双成, 李玉玲, 张云秋. 信息检索课强化思维训练研究. 图书馆学研究 2011, (12): 59.
32. 靖培栋. 信息可视化——情报学研究的新领域. 情报科学, 2003 (7): 685-687.
33. 徐刚, 徐聪. 信息可视化技术在数字化图书馆中的应用. 武汉大学学报(哲学社会科学版), 2008 (6): 919-921.
34. 李学静, 谢蓉. 信息可视化与 Web 信息检索. 图书馆理论与实践, 2004 (3): 94-96.
35. 韩君. 文献信息检索技术的发展. 江苏技术师范学院学报, 2006 (5): 80-83.
36. 胡德华, 张洁, 方平. 生物信息学数据库调查分析及其利用研究. 生物信息学, 2005, (1): 22-25.
37. 贾栋, 贾小云, 马瑞燕. 生物信息学数据库及查询. 山西农业大学学报(自然科学版), 2013, 06: 520-525.
38. 樊文强. 基于关联主义的大规模网络开放课程(MOOC)及其学习支持[J]. 远程教育杂志, 2012, 30 (3): 31-36.
39. Gaines JK, Levy LS, Cogdill KW. Sharing MedlinePlus (R)/MEDLINE (R) for information literacy education (SMILE): a dental public health information project. Med Ref Serv Q. 2011. 30 (4): p. 357-64.

40. Global minimum essential requirements in medical education. [EB\OL]. [2014-10-20]. <http://www.iime.org/documents/gmer.htm>.
41. Information Literacy Competency Standards for Higher Education. [EB\OL]. [2009-06-11]. <http://www.ala.org/ala/acrl/acrlstandards/informationliteracycompetency.htm>.
42. Barclay D. (1993). Evaluating library instruction: Doing the best you can with what you have. RQ, 33, 195-202.
43. International Society for Technology in Education (2000). National educational technology standards for teachers (Online). [EB\OL]. [2009-06-11]. http://cnets.iste.org/teachers/t_stands.html.
44. Medical Library Association. [EB\OL]. [2014-8-31]. <https://www.mlanet.org/>
45. Embedding Informatics in Clinical Education. [EB\OL]. [2014-8-31]. <http://www.eiceresources.org/>.
46. <http://www.oclc.org/dewey/resources/summaries/Deweysummaries.pdf>.
47. <http://www.nlm.nih.gov/class>.
48. <http://www.ihtsdo.org/snomed-ct/>.
49. <http://www.ihtsdo.org/snomed-ct/>.
50. <http://www.nlm.nih.gov/mesh/mbrowser.html>.
51. <http://www.nlm.nih.gov/research/umls/umlsmain.html>
52. <http://news.xinhuanet.com>.
53. <http://www.sinomed.ac.cn>.
54. <http://www.chk.cnki.net/>.
55. <http://med.wanfangdata.com.cn/>.
56. <http://lib.cqvip.com/>.
57. <http://www.clinicalkey.com/>.
58. <http://gateway.ovid.com/>.
59. <http://onlinelibrary.wiley.com/>.
60. <http://www.springer.com>.
61. <http://www.sciencedirect.com>.
62. <http://search.epnet.com>.
63. <http://www.proquest.com>.
64. <http://www.karger.com>.
65. <http://www.ncbi.nlm.nih.gov/genbank/>.
66. <http://www.ebi.ac.uk/ena/>.
67. <http://www.ddbj.nig.ac.jp/>.
68. <http://pir.georgetown.edu/>.
69. <http://www.rcsb.org/pdb>.
70. <http://www.ncbi.nlm.nih.gov/genome>.
71. <http://www.ensembl.org/index.html>.
72. <http://genome.ucsc.edu/>.
73. <http://www.ncbi.nlm.nih.gov/projects/SNP/>.
74. <http://www.omim.org/>.
75. <http://baike.baidu.com/view/798036.htm?fr=aladdin>.

中英文名词对照索引

DIRLINE Directory of Information Resources Online 61
RSS really simple syndication 54
SNOMED CT 模型元数据 SNOMED CT model
component 48
SPIRE spatial paradigm for information retrieval and
exploration 29

B

被引用次数 times cited 140
被引用文献 cited paper 140
本体 ontology 37
标本 specimen 47
标题词语言 heading 36
补充概念 supplement concept 49
布达佩斯开放存取倡议 Budapest Open Access Initiative,
BOAI 245

C

参考文献 reference 140
操作 procedure 46
查找 retrieval 18
超级叙词表 Metathesaurus 51
词干检索 stem search 25
丛书 series 59
存储 storage 18

D

单核苷酸多态性 single nucleotide polymorphisms, SNPs
236
单元词语言 uniterm 36
蛋白质信息资源 the Protein Information Resource, PIR
220
《道兰图解医学词典》 Dorland's Illustrated Medical
Dictionary 61
地址 address 30
电子图书 electronic book 62
杜威十进制分类法 Dewey Decimal Classification 40
短信服务模式 short message service, SMS 56

E

二次文献 secondary document 15

F

发现系统 discovery system 55
非规范语言 uncontrolled language 36
分期与分度 staging and scales 47
副主题词 subheadings 49

G

概念 concepts 45
个人数码代理 personal digital assistant, PDA 56
公共科学图书馆 The Public Library of Science, PLOS 247
关键词语言 keyword 36
观察对象 observable entity 46
光盘 compact disc, CD 18
规范语言 controlled language 35
国际标准连续出版物号 international standard serial number,
ISSN 14
国际标准书号 international standard book number, ISBN
14
国际疾病分类法 International Classification of Diseases,
ICD 42
国际药学文摘 International Pharmaceutical Abstracts,
IPA 193
国际专利分类 International Patent Classification, IPC 169
国家科技图书文献中心 National Science and Technology
Library, NSTL 59
过滤 filters 88

H

核心论文 core papers 267
黑盒子 back box 29
环境或地理位置 environment or geographic location 47
灰色文献 grey literature 16

J

基于草图检索 sketch-based retrieval 27
基于范例检索 example-based retrieval 27
基于内容的检索 content based retrieval, CBR 27
基于内容的图像检索 content based image retrieval,
CBIR 27
基于文献的知识发现 literature based discovery, LBD
256, 270

记录 record 21
 记录人工制品 record artifact 47
 加权检索 weighing searching 26
 检索词典 retrieval thesaurus 34
 检索入口 searching entry 22
 检索史 history 30
 检索语言 retrieval language 34
 健康信息素养 health information literacy, HIL 10
 将信息学嵌入临床教学计划 embedding informatics in
 clinical education, eICE 10
 截词检索 truncation search 25
 精确检索 exact searching 26

K

开放存取 open access, OA 244
 可视化检索 visual retrieval 28
 跨库检索 cross database searching、one-stop searching
 26
 款目词 entry terms 49
 扩展检索 expand searching 26

L

联合专利分类 Cooperative Patent Classification, CPC
 170
 联机公共查询目录 Online Public Access Catalogue,
 OPAC 54
 邻近检索 proximity searching 26
 临床表现 clinical finding 46
 临床查询 Clinical Queries 87
 零次文献 zero document 16

M

美国国家医学图书馆分类法 *National Library of Medicine
 Classification*, NLMC 40
 美国科技信息研究所 Institute of Scientific Information,
 ISI 156
 美国图书馆协会 American Library Association, ALA 6
 模糊检索 fuzzy searching 26
 《默克诊疗手册》*The Merk Manual of Diagnosis and
 Therapy* 61
 目录索引类搜索引擎 search index/directory 240

P

批判性思维 critical thinking 252

Q

期刊引用报告 Journal Citation Reports, JCR 60
 全球医学教育最低基本要求 global minimum essential
 requirement medical education 1
 全文搜索引擎 full text search engine 240

全文型数据库 full-text database 23

R

人工语言 artificial language 35
 人名录 Who's Who 61

S

三次文献 tertiary document 16
 社会环境 social context 47
 身体结构 body structure 46
 失效专利 invalid patent 169
 事件 event 47
 事实数值型数据库 fact-date database 23
 首席参考 First Consult 205
 书目型数据库 bibliographic database 23
 输词查询 Search 50
 树形结构导航查询 Navigate from tree top 50
 树状结构 tree structure 49
 搜索引擎 search engine 240

T

特殊概念 special concept 47
 同被引聚类分析 co-citation cluster analysis 270
 同义词 synonym 46
 同族专利 patent family 169

W

完整的指定名称 fully specified name, FSN 45
 网络信息资源 information resource on network 239
 唯一的标识符 concept ID 45
 文本 text 19
 文本挖掘 text mining 263
 文档 file 21
 文献 literature, document 12
 物理力 physical force 47
 物理性物体 physical object 47
 物质 substance 47

X

限定词 qualifiers 49
 限定检索 limit searching 26
 限定值 qualifier value 47
 限定字段检索 limit field searching 26
 详细病历 situation with explicit context 47
 信息 information 12
 信息共享空间 information commons, IC 58
 信息可视化 information visualization 28
 信息素养 information literacy 3
 叙词 descriptor 36
 学习共享空间 learning commons, LC 58

Y

- 药物/生物产品 pharmaceutical/biologic product 47
一次文献 primary document 15
医学百科全书 Medical Encyclopedia 61, 208
医学图书馆学会 Medical Library Association, MLA 10
医学网景 Medscape 206
引文 citation 30
引文分析 citation analysis 266
引文检索 citation search 140
引用文献 citing paper 140
优先词 preferred term 46
优质信息促进优质卫生保健 better informed for better health and better care 10
有代表性的样本 representative sample 256
有机体 organism 47
语义网络 Semantic Network 52
预印本 preprint 249
元搜索引擎 META search engine 241

Z

- 知识 knowledge 12
知识基础 intellectual base 270
知识型数据库 knowledge database 23
知识转化项目 knowledge translation program 9
中国高等教育文献保障系统 China Academic Library & Information System, CALIS 58

- 中国疾病知识总库 China Disease Knowledge Total Database 23, 203
中国科技期刊开放获取平台 China Open Access Journals, COA 248
中国科学引文数据库 Chinese Science Citation Database, CSCD 149
中国生物医学文献服务系统 简称: SinoMed 66
中国医院知识总库 China Hospital Knowledge Database, CHKD 71
中国引文数据库 Chinese Citation Database 151
中国知识基础设施工程 China National Knowledge Infrastructure, CNKI 72
终端无线通讯协议 Wireless Application Protocol 56
终身教育 lifelong education 3
主题词 subject headings 30, 36
专家词典 SPECIALIST Lexicon 52
专利 patent 15, 168
专利权人 patentee 170
专利申请人 applicants 170
专利受让人 assignee 170
专利优先权 patent priority 169
字段标识符 field tag 22
自动提示 Auto Suggest 91
自然语言 natural language 36
自引 self-citation 140

继承与创新是一本教材不断完善与发展的主旋律。在该版教材付梓之际,我们再次由衷地感谢那些曾经为该书前期的版本作出贡献的作者们,正是他们辛勤的汗水和智慧的结晶为该书的日臻完善奠定了坚实的基础。以下是该书前期的版本及其主要作者:

7年制规划教材

全国高等医药教材建设研究会规划教材

全国高等医药院校教材·供7年制临床医学等专业用

《医学文献检索》(人民卫生出版社,2001)

主 编 方 平

普通高等教育“十五”国家级规划教材

全国高等医药教材建设研究会·卫生部规划教材

全国高等学校教材·供8年制及7年制临床医学等专业用

《医学文献信息检索》(人民卫生出版社,2005)

主 编 方 平

副主编 夏知平

普通高等教育“十一五”国家级规划教材

全国高等医药教材建设研究会规划教材·卫生部规划教材

全国高等学校教材·供8年制及7年制临床医学等专业用

《医学文献信息检索》(第2版,人民卫生出版社,2010)

主 编 罗爱静

副主编 马 路 于双成

编 委 (以姓氏笔画为序)

于双成 (吉林大学)

周旭毓 (中山大学)

于靖涛 (哈尔滨医科大学)

周晓政 (南京医科大学)

马 路 (首都医科大学)

赵文龙 (重庆医科大学)

王虹菲 (大连医科大学)

胡德华 (中南大学)

任慧玲 (协和医科大学)

夏知平 (复旦大学)

何 怡 (天津医科大学)

黄碧云 (中南大学)

张士靖 (华中科技大学)

崔 雷 (中国医科大学)

罗爱静 (中南大学)

谢志耘 (北京大学)

金龙玉 (中南大学)

学术秘书 谢文照 (中南大学) 龚佳剑 (首都医科大学)

医学文献信息检索

Medical Literature Information Retrieval

第3版

1. 细胞生物学 (第3版) Cell Biology
2. 系统解剖学 (第3版) Systematic Anatomy
3. 局部解剖学 (第3版) Topographic Anatomy
4. 组织学与胚胎学 (第3版) Histology and Embryology
5. 生物化学与分子生物学 (第3版) Biochemistry and Molecular Biology
6. 生理学 (第3版) Physiology
7. 医学微生物学 (第3版) Medical Microbiology
8. 人体寄生虫学 (第3版) Human Parasitology
9. 医学遗传学 (第3版) Medical Genetics
10. 医学免疫学 (第3版) Medical Immunology
11. 病理学 (第3版) Pathology
12. 病理生理学 (第3版) Pathophysiology
13. 药理学 (第3版) Pharmacology
14. 临床诊断学 (第3版) Clinical Diagnostics
15. 实验诊断学 (第3版) Laboratory Diagnostics
16. 医学影像学 (第3版) Medical Imaging
17. 内科学 (第3版) Medicine
18. 外科学 (第3版) Surgery
19. 妇产科学 (第3版) Obstetrics and Gynecology
20. 儿科学 (第3版) Pediatrics
21. 感染病学 (第3版) Infectious Diseases
22. 神经病学 (第3版) Neurology
23. 精神病学 (第3版) Psychiatry
24. 眼科学 (第3版) Ophthalmology
25. 耳鼻咽喉头颈外科学 (第3版) Otorhinolaryngology-Head and Neck Surgery
26. 核医学 (第3版) Nuclear Medicine
27. 预防医学 (第3版) Preventive Medicine
28. 医学心理学 (第3版) Medical Psychology
29. 医学统计学 (第3版) Medical Statistics
30. 循证医学 (第3版) Evidence-based Medicine
31. 医学文献信息检索 (第3版) Medical Literature Information Retrieval
32. 临床流行病学 (第2版) Clinical Epidemiology
33. 肿瘤学 (第2版) Oncology
34. 生物信息学 (第2版) Bioinformatics
35. 实验动物学 (第2版) Laboratory Animal Science
36. 医学科学研究导论 (第2版) Introduction to Medical Research
37. 医学伦理学 (第2版) Medical Ethics
38. 皮肤性病学 Dermatology and Venereology

策划编辑 / 常 帆

责任编辑 / 常 帆 崔曼曼

封面设计 / 李 蹊

版式设计 / 李秋斋

本书附赠网络增值服务, 激活方法:

1. 注册并登录人卫医学网教育频道 (edu.ipmph.com)
2. 点击进入“网络增值服务”, 搜索找到本书
3. 点击“激活”并输入“激活码”



ISBN 978-7-117-20571-9



9 787117 205719 >

定 价: 52.00 元

[General Information]

书名=医学文献信息检索

作者=罗爱静, 于双成主编

页数=295

SS号=13924388

DX号=

出版日期=2015.05

出版社=北京人民卫生出版社