

现代信息检索

主 编 郑章飞

副主编 敬 卿 陆 璇

华中理工大学出版社

图书在版编目(CIP)数据

现代信息检索/郑章飞 主编 敬卿 陆璇 副主编
武汉:华中理工大学出版社, 1999年8月
ISBN 7-5609-1994-4

I. 现…

I. ①郑… ②敬… ③陆…

II. 情报检索-高等学校-教材

N. G354

现代信息检索

主编 郑章飞
副主编 敬卿 陆璇

责任编辑:李丛晖

监 印:张正林

出版发行:华中理工大学出版社

武昌喻家山 邮编:430074 电话:(027)87542624

经销:新华书店湖北发行所

录排:华中理工大学出版社照排室

印刷:武汉市新华印刷厂

开本:850×1168 1/32

印张:15

字数:360 000

版次:1999年8月第1版

印次:1999年8月第1次印刷

印数:1—10 000

ISBN 7-5609-1994-4/G·247

定价:16.80元

(本书若有印装质量问题,请向出版社发行科调换)

目 录

基础知识篇

第一章 绪论	(2)
1.1 现代信息系统的五元组成	(2)
1.2 信息交流系统	(4)
1.3 信息及信息工作的作用	(5)
1.4 现代信息技术概说	(7)
1.5 信息检索定义新说	(15)
1.6 信息检索的理论基础	(19)
1.7 信息检索的意义	(26)
第二章 文献知识	(28)
2.1 文献与信息	(28)
2.2 科技文献的出版类型	(29)
2.3 科技文献的级别	(32)
2.4 科技文献的载体形式	(34)
2.5 检索工具及检索方法	(35)
2.6 检索途径与检索步骤	(37)
2.7 缩略语与音译	(40)
2.8 文献情报调研	(46)
第三章 检索语言	(49)
3.1 检索语言的基础	(49)
3.2 体系分类法	(52)
3.3 标题词语言	(54)

3.4 关键词语言	(56)
3.5 叙词语言	(58)

科技信息检索篇

第四章 中文科技期刊文献的检索	(64)
4.1 《全国报刊索引》(自然科学与技术版)	(64)
4.2 《中国机械工程文摘》	(67)
4.3 理工类中文科技检索刊物选录	(68)
第五章 外文科技期刊文献的检索	(70)
5.1 《工程索引》	(70)
5.2 《科学文摘》	(76)
5.3 《金属文摘》	(82)
5.4 《化学文摘》	(90)
5.5 《科学引文索引》	(100)
第六章 专利文献的检索	(108)
6.1 知识产权与专利文献	(108)
6.2 中国专利文献的检索	(113)
6.3 美国专利文献的检索	(116)
6.4 《世界专利索引》	(120)
第七章 科技报告的检索	(127)
7.1 科技报告的类型和特点	(127)
7.2 中国科技报告的检索	(128)
7.3 美国四大报告的检索	(129)
第八章 会议文献的检索	(137)
8.1 概述	(137)
8.2 会前出版物的利用	(138)
8.3 会后出版物的查找	(143)
第九章 其他类型文献的检索	(150)

9.1 书籍的检索	(150)
9.2 标准文献的检索	(151)
9.3 学位论文的检索	(157)
9.4 产品资料的检索	(161)
9.5 科技专业名词术语的检索	(166)
9.6 技术数据的检索	(169)
9.7 科技百科知识的检索	(174)
9.8 科技人物资料 and 组织机构的检索	(176)

社科信息检索篇

第十章 工具书概述	(180)
10.1 什么是工具书	(180)
10.2 工具书的类型	(180)
10.3 工具书的编排与查检	(188)
第十一章 各类社科文献的检索	(193)
11.1 书籍的检索	(193)
11.2 报刊资料的检索	(198)
11.3 字、词的检索	(202)
11.4 专科词语和百科知识的检索	(206)
11.5 年、月、日的检索	(213)
11.6 人物资料 and 组织机构的检索	(223)
11.7 地名、地理资料的检索	(229)
11.8 历史事件、当代大事的检索	(233)
11.9 法规和统计资料的检索	(238)

计算机信息检索篇

第十二章 计算机检索基础知识	(244)
-----------------------------	-------

12.1	计算机检索的发展	(244)
12.2	计算机检索系统的组成及服务	(249)
12.3	数据库及其编排结构	(250)
12.4	计算机检索方法	(256)
第十三章	联机检索	(269)
13.1	国际联机检索	(269)
13.2	国内联机检索	(285)
第十四章	光盘检索	(288)
14.1	《工程索引》光盘检索方法	(288)
14.2	美国政府科技报告光盘检索方法	(301)
14.3	《化学文摘》光盘检索方法	(310)
14.4	《INSPEC》光盘检索方法	(315)
14.5	《中文科技期刊数据库》光盘检索方法	(322)
14.6	《中国学术期刊(光盘版)》检索方法	(330)
14.7	万方数据库系列光盘及其检索方法	(340)
14.8	其他常用中文数据库光盘介绍	(348)
第十五章	Internet 网上信息检索	(360)
15.1	Internet 概述	(360)
15.2	Internet 搜索引擎	(363)
15.3	Internet 网上各类信息的检索	(378)
15.4	CALIS 服务功能及其提供的信息资源	(387)

信息利用篇

第十六章	科技论文写作与投稿	(396)
16.1	什么是科技论文	(396)
16.2	如何选题	(397)
16.3	稿面形式和写作方法	(406)
16.4	科技论文投稿知识	(415)

第十七章 科学研究与信息资源	(436)
17.1 科学研究中的文献信息资源保证	(436)
17.2 信息资源建设与开发利用	(441)
附录 1 Internet 常用搜索引擎选录	(455)
附录 2 图书馆网址选录	(458)
附录 3 其他相关网址选录	(464)
参考文献	(466)

基础知识篇

第一章 绪 论

1.1 现代信息系统的五元组成

当今,世界各国都在搞“信息战”,对情报信息的竞争十分激烈。在1996年8月26日,中央电视台《焦点访谈》节目“看不见的战线”里报导:“美国驻中国使馆大使说:‘现在是获取中国情报的黄金时期。’”我国著名学者钱学森在《经济生活报》中提出:把情报信息作为第四产业,它生产非物质形态的思想和各种信息符号,必须引起高度重视,科学工作者应具备熟练的检索技能,以适应信息时代的要求。

现代信息系统由5个部分组成,即:信息、计算机、网络、信息库和人。下面我们一一加以说明。

1.1.1 信息

信息可以是人脑中的信息,也可以是自然界或人造系统中的信息。著名控制论专家维纳(Wiener)在他的《信息控制论》中说:“信息是人们在适应外部世界并使这种适应反作用于外部世界过程中,同外部世界进行交换的内容的总称。”正是信息构成高级社会的一种资源,是物质、能源以外的第三种资源。它就像货币一样,流动起来可以增值,产生财富,推动社会进步。

1.1.2 计算机(电脑)

它是一种当代社会不可缺少的信息处理工具。计算机网络可以由巨型机、大型机、中型机、小型机和微机组成。微机是家庭或办

公室中的节点机;中小型机是具有各种数据库、知识库,能作复杂处理的节点机;大、中、巨型机是并行和共享的。进一步发展,将是智能电脑。智能电脑系统有丰富的系统软件和各种应用软件、智能软件、学习软件等作为共享资源。

1.1.3 互联网络

因特网就是互联网络的一种形式。这种网中网,可以包容各种分布性网、计算机集成制造系统网和大众媒体传播系统网等。在这种网络中,任何人、任何时间都能与任何地方的另一个人进行通信,所提供的业务仅受接入网或终点网及用户终端能力的限制。互联网最终将能向用户提供任何形式的信息服务,从而真正实现“秀才不出门,能知天下事”的梦想。

1.1.4 信息库

计算机和网络能发挥巨大作用的基础之一是信息库。其作用的发挥,与库内信息的优质和严密的组织结构、信息库的组织情况及数据存取效率关系极大,这尤其表现在各种关系错综复杂、交叉引用的多媒体环境下。互联网是个非常复杂的巨系统,在系统软件的配合下,它能自动建库,保持库的活力,给问题求解以巨大的信息支持。

1.1.5 人

这是任何系统的主导,是系统智力的源泉。人机联作能形成各种各样的互补。例如,目前智能系统知识少,遇到尴尬局面就需请人类专家帮助;机器可做比人效率高的工作,如控制高速运行的机器,在高温、低温、有毒、有危险的环境下工作。总之,任何智能系统都离不开社会性、人文性。系统是人设计的,用户是人,处理的对象与人类息息相关。人要在系统中发挥能动作用,还需注意人机工效学的理论,以提高人机效率;人还得安全避害、舒适;系统要降低消

耗。

1.2 信息交流系统

当代经济和社会的发展,越来越依赖于科学和技术,而科学的发现和技术发明又取决于科学家个人或科学研究集体吸收前人或他人的研究成果的能力。现代科学研究都是一种具有广泛社会联系的活动,其进步是通过集体的努力而获得的,并且是一种循序渐进的过程。前人的成果靠后人继承发展,一个集体的工作建立在另一个集体的工作基础之上。对于所完成的研究成果,只有在得到合适的表述和有效的传播交流后,才能为人们所吸收利用,才能体现出成果的真正价值。科学成果如果得不到有效的传播交流和利用,科学就要停滞。离开有特色的信息传播交流方式,科学的形成和发展也是不可能的。

所谓信息交流,就是人们借助于共同的符号系统进行知识的有效传递。它包括正式交流过程和非正式交流过程。

1.2.1 非正式交流过程

这是由科研人员及工程技术人员自己直接完成信息交流的过程。包括:

- (1) 科学技术工作者之间就他们从事的研究和研制进行直接对话,如交谈、参加学术会议等;
- (2) 科技工作者参观同行的实验室、科技展览等;
- (3) 演讲、书信往来等。

1.2.2 正式交流过程

主要由文献信息人员借助于科学文献系统进行交流的过程。包括文献编辑、出版、发行和科学信息工作。

综上所述,非正式交流过程更多地表现为信息创造者和信息

用户个人的行为特征。而正式交流过程则表现为文献交流和传播，文献检索就是其中非常重要的一环。

1.3 信息及信息工作的作用

由于现代科学技术的高度发展，科学文献数量剧增，学科众多，彼此交叉渗透，综合化程度愈来愈高，科学技术对社会生产的影响也越来越大，因而对情报工作提出了更多更高的要求。科学信息工作(包括文献检索工作在内)对发展科学技术和社会生产的重大意义和作用，受到普遍重视。目前国内外流传较广的所谓信息工作三论(资源论、经济论、决定论)，比较集中地反映了人们对信息工作重要性的评价，具有代表性。

1.3.1 资源论

资源论认为信息是一种重要的资源，是人类社会的“第二资源”，是构成生产力的最活跃的因素。善于开发利用信息这种资源，就能有效地促进经济和社会的发展。许多国家都从实践中认识到了开发利用信息资源的重要性。在这方面，日本的例子是比较突出的。我们都知道，日本是一个地少人多，物质资源贫乏的国家。但是战后日本的经济增长速度快得惊人，在短短的几十年间，日本一跃成为世界经济大国。美国哈佛大学教授、美日关系研究室主任埃兹拉·弗格尔写了一本叫做《日本——世界第一》的书，书中有这样一段话：“如果只举一个原因说明日本的成功，那就是对知识信息不间断地、集体地追求。日本的领导者和人民都把学习和追求知识信息看做毕生最重要的事情。”这就是说，日本人善于利用知识资源。如战后日本的石油化工、电子等新兴工业的新技术几乎全盘来自国外，都是利用国外信息资料搞起来的。有些技术如数控机床技术，则是在国外已经接近完成的情况下，日本信息人员搞来了资料，然后日本科学家在别人基础上很快就研制成功了。资源论的观

点在日本非常流行。

1.3.2 经济论

经济论认为无论是搞科研,还是发展社会生产,都只有通过信息工作,在他人最新经验及成果的基础上去进行,才是最有效最经济的。那种闭关自守,一切都靠自己单干,重复别人劳动的科研,不仅不会有高速度、高水平,而且还会造成人力、物力和时间的极大浪费,这已为大量的实践所证实;反之,一切都依赖外国,靠进口高价成套技术设备来发展工业,既不合算,其外汇支付也是一般发展中国家所难以承担的。只有充分发挥信息的作用,有选择地引进某些必要的适用技术,并从实际需要出发,有计划地确定那些更具迫切性的科研项目,进而发展本国的科学技术研究体系,才是最经济可靠的。这就是所谓:“科学研究是高价的,科研成果是昂贵的,科学技术信息则是便宜的。”根据有关部门统计,科研人员查阅所需的信息资料,一般得占用全部研究工作时间的 1/3 左右。在我国,由于信息不灵,一些科研部门的科研项目据估计约有 40% 和国外同行重复,是别人已经取得了成果的。这应该说是一种极大的浪费。有人作过这样的估计,假如有 100 万科技人员,按照过去的科研办法,就等于有 30 多万人长期在做查找信息资料的工作。如果信息服务部门能帮助他们,即使是将 30% 的查找资料时间减少为 10%,这就相当于增加了 20 万科技人员;如果再将上述由于信息不灵造成的重复科研项目减少到最低限度,则相当于又增加了一批科研力量。对我们这样一个科学技术比较落后,科研力量深感不足的发展中国家来说,这是不能不算的一笔大账。

1.3.3 决定论

随着现代科学技术的迅速发展,欧美各国的工业技术基础日益雄厚,经济能力不断增长,在科学技术的许多领域,它们的水平不相上下,彼此竞争十分激烈。可以说,谁高明一点,谁的信息灵,

谁就能在国际贸易市场上压倒对方;或者说,谁善于开发利用信息,特别是善于从对方搞到有关技术的最新信息,谁就能在科技方面迅速赶上甚至超过对方。在这种情况下,信息资料往往起着决定性的作用。这就是决定论的观点。

1.4 现代信息技术概说

1.4.1 计算机科学技术

我们在讨论人工智能和信息检索的关系时,无论从哪一方面,似乎都应该谈谈计算机技术,因为它不仅是双方使用的工具,而且是维系当代社会的“金三角”(即计算机、数据库和人类智能)中的一角。数据库(或知识库)是人类几千年智慧的结晶,计算机系统(包括人工智能系统)是人类智能的延伸,将来可能作为人的伙伴与人“平起平坐”(当然它还是受人支配的人造物),人类智能是“金三角”中的控制因素,无疑最重要。当代世界三元素之间由人造的通信网联起来,成为信息的有机的整体系统。

1. 计算机的历史

计算机的历史可简单地分为两个阶段:冯·诺依曼机阶段和非冯·诺依曼机(智能机)阶段。在冯·诺依曼机阶段(1946—1982),从1954年起始的用计算机处理文字,是计算机应用史上的里程碑事件。60年代后的计算机与通信网的结合具有非常重大的意义。发展到今天的开放系统互联网,加上计算机会议、多媒体技术,给历史上的文化积淀与现代人头脑中的知识交流创造了无与伦比的科技环境,改变了人类研究、设计、决策、工作和思维的方式,为人类高效工作立下了不朽之功。1981年,日本宣布的第五代机研制计划虽未成功,但对智能机的研制工作起了极大的推动作用。智能机的研究,受无须说明的智能转移规律(辅人率、拟人率与人机共生率)影响,早晚必将获得成功。

2. 计算机、人工智能和信息检索

计算机是一种信息处理工具,这是当初研究和创造计算机的人所没有想到的。人工智能、信息检索与计算机结合,借助计算机来揭示人类信息加工的奥秘,并进而把不胜其烦的信息处理转移到智能机系统上。这在计算机与人工智能、信息检索的关系上,真可以说是“一拍即合”。就目前人脑研究的成就看,人脑的主要功能就是信息加工。计算机目前的水平也只能做信息加工工作。所以说,计算机找到了用武之地,人工智能和信息检索找到了最合适的工具。

3. 未来的计算机

未来的计算机一定是智能机。有文献指出未来智能机的几大功能:能存储大量信息和知识;会思维和推理;能进行高速并行处理;能用自然语言、文字或图像等与人进行信息和知识的交流;能试探求解问题和有一定创造力;具有学习功能和能不断修正、完善自身;能解决多种问题和只需告诉它作什么,就可以自动工作。图 1-1 是一种智能机示意图。除了创造力外,可以认为,智能机就是知识处理系统,与信息检索系统同质、同构。

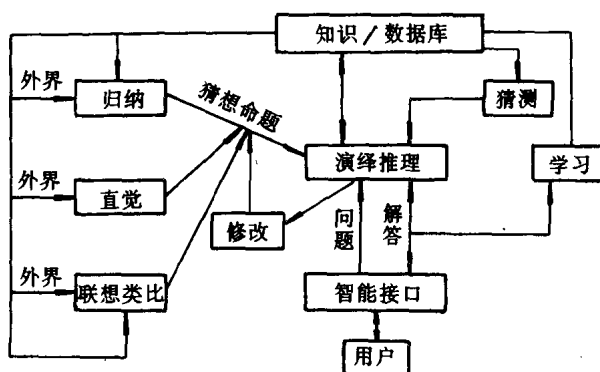


图 1-1 智能机示意图

1.4.2 现代通信技术

1. 概论

在信息已成为最重要的战略资源的今天,如果说计算机系统就是现代社会的大脑,那么通信系统就是现代社会的神经系统。计算机与通信结合,才能完成现代信息的战略资源使命。

现代通信可以认为是由各种媒体表达组成的,这些表达对通信双方共有的一个知识结构产生激发,从而完成信息传递。现代通信的功能如图 1-2 所示。

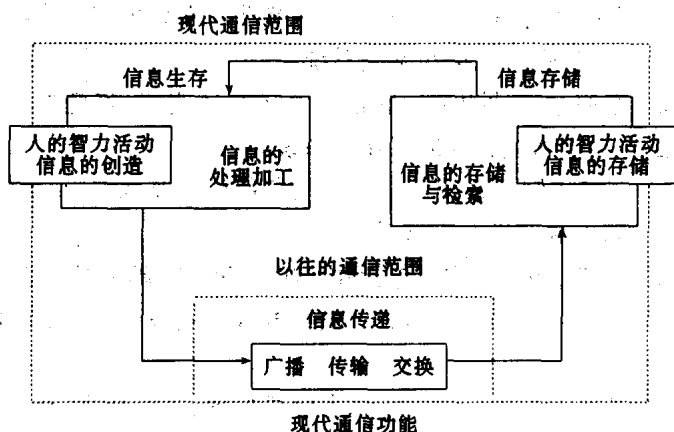


图 1-2 现代通信的功能示意图

现代通信的介质有声、光、电、磁等。通信技术有卫星通信和光纤通信等。通信的系统可有公用系统、事物系统和家庭系统。这些系统又通过网络联在一起。网络有局域网、远程网,这些网络又通过协议,形成开放系统互联网,如 Internet 网。

2. 家庭系统

现代社会的许多活动都在家庭中进行。下面,我们就以家庭系统为例,描述一下通信情况。

由于家庭办公及生活、娱乐信息的处理,家庭系统是未来通信

系统网中的一个极为重要的终端系统。由于人类社会正进入信息时代,暂时还不能对家庭系统进行具体的描述,但大的框架一般可能如下:

(1) 家庭系统的信息。家庭系统的信息主要有 4 类:

(a) 娱乐信息。家庭是人们休息的地方,各种娱乐的信息将是最主要的、大量的信息。

(b) 生活信息。如新闻、天气预报、经济及行情、天灾等。

(c) 生活行动信息。如购物、利用金融机构办理各种手续,预订购车、船、机票等。

(d) 家庭办公信息。由于创造性劳动和家庭主妇工作要求的需要,办公信息也将包括在家庭系统中。

(2) 家庭系统的组成。它由 3 个子系统组成:家庭控制子系统、家庭生活子系统和家庭工作子系统。

(a) 家庭控制子系统。指居室内各部位的监视器、冷暖气、门窗的开闭和电源切换及报警系统。

(b) 家庭生活子系统。包括收音机、电视机、新功能电话、可视数据(如用户智能电报)、CATV(有线电视)、业务咨询、家庭购物、存款。

(c) 家庭工作子系统。它具有信息处理、文字处理、访问企业信息处理中心、记录通信的功能。

未来的家庭系统中,将有以家庭计算机为中心的各种信息业务终端,并与家庭网络(点线)相联,然后接入各种局域网。

(3) 家庭系统及业务动向。家庭系统的设备主要是电视机、电话、个人计算机、收录机等。这些设备组合在一起就变成各种媒体终端,其中计算机将管理出入家庭和家庭内的信息。它们有加工处理、存储检索及文字处理功能。各种计算机将分别适合家庭不同层次成员的选择,如家庭办公用的事务处理机、处理家务用的计算机和学习用的儿童机等。

(4) 家庭系统用的接入系统。接入家庭系统的通信系统包括: