

综述·评介

●张帆 娄策群

迈向 21 世纪的检索技术

ABSTRACT Being a basic means for people to store in and access to informations, information retrieval, among other subjects of library and information sciences, is a branch of knowledge with a maximum of vitality. On the basis of evaluating and assessing the existing state of the development of computer search and data base techniques, the paper forecasts the theoretical research and technological development of information retrieval. 9 refs.

SUBJECT TERMS Retrieval techniques - 21st century Information retrieval - Summaries

CLASS NUMBER G354

情报检索是人类有目的存贮和查找信息的基本手段,是图书情报事业赖以生存和发展的基础,也是图书馆学情报学中最具活力的分支学科。近 20 年来,在新技术革命浪潮的冲击下,人类开发和利用信息资源的环境发生了前所未有的变化。早在 70 年代,国外便有人预言:电子计算机、光盘技术和光纤通讯等新技术问世及其相互结合将引起情报检索技术的革命,其软件也将发生根本性变革;作为新型的智能系统而存在的情报检索系统将不仅能够识别声音、图像、处理自然语言,而且将作为专家系统和知识库来帮助人们生产和利用知识,处理和解决各种问题。目前,情报检索正从文献机构的深宅大院走向市场,走向社会大舞台。特别引人注目的是,已逐渐形成一个以缩微技术、光盘技术、数据库

技术为主体的现代情报检索高新技术群,而且这些技术相互结合将为情报检索现代化开辟更加广阔的道路。

1 计算机检索

现代情报检索以电子计算机在情报检索中的应用为标志。从世界上第 1 台电子计算机诞生至今,计算机检索已经历 50 年代早期的单元词组配检索、60 年代脱机批式检索和联机检索试验以及 70 年代联机网络检索 3 个阶段。

我国计算机检索及其系统建设虽然起步较晚,但发展日新月异,目前已达到相当高的水平。从总的方面看,我国计算机检索系统的建设大体经过了以下过程。

1.1 引进国外磁带,开展定题服务(SDI)

我国计算机检索最早由商业途径和交换方式从国外引进文献磁带,进行定题检索和联机检索开始。据统计,1975年我国首次引进国外磁带,到1992年底,中国科技信息所,机械科技情报所,北京文献服务处和上海科技情报所等14个单位先后引进文献磁带65种,自建数据库36个,总容量达1700万条,服务课题6万余个。其中有7个单位建有国内联机终端600余个,主要分布在北京、上海、武汉和广州等地。引进的磁带主要有《化学文摘》(CA)、《美国政府报告》(GRA)、《工程索引》(COMPENDEX)和《世界专利索引》(WPI)等国际著名磁带。这些磁带的引入在一定程度上满足了我国科技人员检索信息的需要,同时也为我们吸收国外已有软件技术、创建汉语数据库提供了借鉴。

1.2 汉字信息处理与情报检索软件设计

汉字信息处理与汉字软件研制是计算机信息管理领域的重要课题,它涉及信息学、计算机科学、语言学和系统工程等多种科技知识。80年代以来,我国汉字信息处理与检索软件研究取得了一些进展,特别是关于汉字输入技术、汉字切分、汉语文献自动标引、自动翻译和软件建设等方面的研究已由试验进入发展阶段。据初步统计,我国已经研究出约500种汉字输入方案,一批兼具汉字处理、编辑、检索和管理功能的软件正处在完善和推广使用之中,各具特色的自动切词、自动标引方案正脱颖而出。其中王永成和吴蔚天的“部件词典法”和“后缀表法”已经过数以万篇文献处理实践,其抽词结果与手工标引效果相差无几。北京大学的“主题词典法”试验兼具自动抽词功能和自动标引特点。与此同时,我国一些图书情报部门一方面对引进的各种磁带进行开发建库,同时在积极研制检索软件方面也取得了一定的成绩。如1987年仅在北京召开的第7次全国计算机情报检索研讨会上,就有34项汉字情报检索与管理系统在大

会上推广。这些试验性工作不仅展示了我国研制情报检索与信息管理软件的实力,也为我国自行研制情报检索软件 and 在国外应用软件的基础上进行二次开发和汉化做了充分的理论与技术准备。现在,我国除广泛使用汉化的国外应用软件CDS/ISIS之外,还在汉化的DBASE、ORACLE数据库管理系统上设计和研制情报检索和其它应用软件。近年来,国内出现的上百种软件几乎都是我国自行研制或在国外应用软件基础上进行二次开发和汉化的产物。如北京文献服务处在UNIVAC 1100/70上设计的中西文联机情报检索系统BDSIRS,不仅具有通用检索功能,而且外存空间利用率较大,体现了我国外存条件的特点。目前,我国自行研制的情报检索软件大都属于面向本系统本部门的封闭型系统,虽然基本上都具有编制词表、建库、检索、编辑、系统管理的综合情报处理功能,但这些软件的设计仍属于初级阶段,即以实际面向规范语言的倒排检索方式和命令驱动用户接口为主要特点,相互不能互换和兼容。对于新型的全文文本检索、用户接口、自然语言 and 标准词的互换等课题尚处于摸索阶段。

1.3 国际联机检索

1980年4月,我国建筑技术开发中心联合9个部委在香港租用1台终端机,首次与美国的DIALOG、ORBIT系统联机,迈出了我国国际联机检索的第1步。1981年5月,石油部科技情报所在北京电传机上与上述两个系统联机。同年12月,北京科技情报所改造了西门子的电传机并与DIALOG系统联机。1983年9月,中国科技信息所在邮电部支持下,建了第1台终端机,经卫星信道接到意大利分组交换中心,与欧洲的EURONET、美国的TYMNET、TELENET网络相连,并通过ESA-IRS系统与美国的DIALOG和ORBIT系统联机。1987年,中科院文献情报中心与西德的FIZKA合作连接STN系统,1989年又与瑞士的DATA-

STAR 系统联通。这样,我国从 1983 年开始建立国际联机检索终端至 1992 年底,已有国际联机终端 130 多个,分布于全国 50 多个城市,引进国外联机系统 20 个,可使用的数据库约 800 个,文献量超过 2 亿篇。

与此同时,国内对社会开放的联机服务也有较快的发展,如中国科技信息所的 ISTICORS 系统和北京文献服务处的 BDSIRS 系统等,都是面向全国的联机服务系统。据统计,还有面向地区或系统的检索终端约 50 个,可供检索的数据库容量超过 400 余万篇。此外,“七五”期间,国家科委信息司重点支持的联机科技情报检索系统实现了 ISTIC、CHOICE、MEIRS 3 家系统 4 台主机在 CNPAC 上联网,北京文献服务处与航天工业部信息网实现了同类主机之间在专用通信网上联网,把联机服务推进了一步。在此,特别值得一提的是,近年来,随着我国市场经济体系的建立,全国数以千计的面向经济、技术、市场和社会的行业、专业、地区的信息系统崛起,一批民间的或企业的联机服务机构相继涌现,这些检索系统一般都采用微机系统建立各种面向市场和用户需求的数据库,通过电话网和电传网向用户提供服务。这种面向市场需求而建立的检索系统,虽然目前条件较差、业务尚不稳定,但其社会化、专业化、产业化程度普遍比政府部门的信息系统高,这是一个充满生机与活力的信息检索领域,应引起政府和社会的广泛关注、扶持和管理。

1.4 微机检索

80 年代以来,微机以其体积小、价格低廉和使用方便等优势而广泛应用于图书情报领域。微机的普遍应用还应归功于其运行环境不受限制、性能不断提高,既可作为联机检索的智能终端,便于用户与系统联机 and 套录数据,又可作为不同系统之间的转换设备或作为建立个人检索系统的工具。据统计,目前我国已有微机 15 万余台,而且还在不断增

长。近年来,微机正以异军突起的巨大潜力和极高的效率冲击、影响着情报检索方式,使情报检索技术进入了一个崭新的发展时期。特别是近几年推出的光学存贮器(光盘),由于其具有海量存贮、成本低廉、寿命长以及使用方便等优点,被誉为继 IC 存贮器之后的第二次存贮器革命。“微机+光盘”已显示出强劲的生命力,正在改变着信息处理和传递的模式与方法。利用光盘在微机上检索,可以不受通信费用的限制,使从大型数据库中套录信息再建数据库成为可能。目前,利用美国 DIALOG 仿真系统开发的光盘套录建库与检索系统 DCRS 所建立的几个含有上千篇记录的专题数据库运转正常,其试验亦达到预定要求。

以光盘取代磁盘而作为微机上的主要存贮器是大势所趋。它将为微机在情报检索方面的广泛应用展示出美好的前景。与此同时,也将为我国情报检索的理论与技术发展提出更多、更新的研究课题。

2 数据库

我国数据库建设同世界上发达国家相比,虽然起步较晚,但发展很快,现已成为我国计算机检索中蒸蒸日上的一个领域。从 1980 年 8 月首届全国计算机检索研讨会至 1991 年底,在国家科委科技信息司登记注册的各类数据库已有 805 个,其中科技/工程数据库最多,占 45%,文教卫生类数据库占 29%,经济贸易类数据库占 16%。目前,文献数据库已趋饱和,而事实数据库和数值数据库正在增长,图象数据库越来越受到重视,全文数据库无论从数量还是质量上都将有较大的发展。

目前,我国建库工作主要由中央各部委和地方所属的情报所承担。图书馆和档案系统也建有一些数据库,记录量超过万条的中国信息源数据库约有 50 多个,其中英文版的

几个数据库已开始进入国际联机检索系统,利用国外磁带建立的数据库有20多个,部分可以提供联机检索服务。在50个城市设立的百余个联机检索终端可以连通国际上12个大型联机检索系统,国家投资建设的公用分组数据交换网(CNPAC)业已开通。另一方面,国产数据库开始打入国际市场,参与国际交流。据悉,目前约有10个国外数据库已将我国相关专业的文摘译成英语,纳入其数据库。中国大学学报论文英文文摘数据库将在美国DIALOG系统中向全世界提供服务。

2.1 多媒体(Multimedia)数据库

多媒体是一个在文本基础上发展起来的概念,它并非媒体的结合,而是将有媒体特征的文本、声音和图形等信息加以组织而形成的远程通讯检索系统。多媒体系统功能强大,需配置多种外围设备来支持信息的输入输出。理想的多媒体给人回归自然的感觉:用户既能看到立体彩色画面,又能置身于有声的环境中,同时还有信息说明。

目前比较流行的多媒体主要是视屏文本(Videotext),它能同时传送文本和图形信息,采用的通讯方式是电话网。视屏文本在法国应用较多。近两年,英国、日本又推出了照相视屏文本(Photo-Videotext)。多媒体技术用于情报检索还处在实验阶段,但代表着现代情报技术的最新成果和未来检索技术的方向。

2.2 全文数据库

全文数据库是一种存贮文献全文或文献中主要部分的源数据库,其检索理论诞生于50年代末。1973年,美国Mead Date Control公司开发的法律全文数据库LEXIS投入使用,开创了全文数据库服务的先河。随着电子排版技术的发展和大量存贮媒体的普及,由电子文档直接形成全文数据库并提供联机检索的系统日益增多。以欧美商用数据库名录为例,1980年版所录的500种数据库中,有25种全文数据库,占5%,而1989年版收

录有1391种全文数据库,已占34%。

国内第1部全文数据库《国共两党关系通史》于1990年7月研制成功,约150万字,在国内出版界引起很大反响。武汉大学全文数据库研究课题组经过4年努力,于1991年在微机上研究成功中西文兼容的“湖北省地方志全文检索系统”,该系统建立了约10万字的《湖北省地方大事记》和《中国人民解放军大事记》全文数据库,为检索湖北地方志提供了源文献。

2.3 超文本数据库

超文本是在文本、图象、声音等文献信息数据基础上建立起来的一种非线性组织形式的数据库。在这种系统中,文献信息单元不是按照线性序列排列,而是依赖数据库中文献单元之间知识内容的内在联系和可能出现的明显的连续性来组织。

在超文本数据中,最小的知识单元为结合点(node),知识单元之间以链路(line)相连接,用户界面则以窗口形式出现,窗口采用图形标记,用户可以通过鼠标器激活图标中的相应结点,以任意顺序通过链路阅读所有的相关资料。超文本的基本特点是既可以将文档及其参考文档或注释等相关信息结点连接起来,也可以将同类文档结点连接起来,从而构成一个不遵循某种正规模式的框架结构。这样,将有助于从事研究活动的专家学者的理解、记忆和想象。

超文本是一种新型的数据库系统,它以联想方式组织情报,以非线性方式进行检索,可用来改善人机界面,提高检索效果。除了文本信息的存取之外,它还可以把图象和声音等信息存取于结点内,同时还可为用户提供灵活多样的查找方法。现在,超文本技术已对传统情报检索技术产生深远影响,研究与发展超文本技术已成为推动情报检索现代化的战略措施。

2.4 CD-ROM 数据库

随着微电子技术的进步,数据库的存贮

介质也在不断增多。自 1985 年世界上第 1 个 CD-ROM 数据库 BIBLIOFILE 问世以来, CD-ROM 发展非常迅速,应用日益广泛。这种数据库通过 CD-ROM 驱动器与 IBM-PC 等微机相联接,可以形成具有相当于联机检索功能的文献检索系统。由于光盘具有海量存贮、寿命长、信息更新周期短等优点,已成为当前一种最新型的情报载体。在我国,国家海洋局情报所于 1986 年最早引进 CD-ROM 光盘。截至 1991 年底,全国约有 CD-ROM 用户 131 户,共引进 57 类 201 种 CD-ROM 数据库。其服务方式主要有:图书订购、图书定题服务、回溯服务、咨询服务、建立专题数据库、高校机检课程教学、机检用户教育与培训、国际联机前的预检等。

在发达国家,CD-ROM 光盘检索系统只是联机检索的一种支持与补充。在今后几年中,我国发展 CD-ROM 的策略应该是提高光盘利用率,尽早实现光盘技术与通信技术的整合,使多用户共享;建立地区、部门、跨行业的服务网,有计划地建立 CD-ROM 阅览室,让用户在显示屏前浏览电子出版物;同时做好 CD-ROM 数据库的引进管理与开发工作,处理好国际联机、光盘、磁带与手工检索的关系,生产国产 CD-ROM 数据库产品,建立我国的光盘产业,发展多光盘系统。

3 21 世纪情报检索理论研究与技术发展展望

现代信息技术的突飞猛进、社会信息化程度的提高和我国经济建设的发展,促使情报检索的理论研究与发展向着“六化”方向发展。

3.1 情报检索高新技术集成化

伴随着现代信息技术的高速发展,信息介质与传递费用的降低,微电子技术、高密度存贮技术、计算机技术和网络通讯技术等将会在情报检索中得到更广泛的应用。情报检

索系统将是现代高新信息技术之集大成者。

3.2 情报检索内容综合化

21 世纪的情报检索将完全突破书目型科技情报检索的范畴。数据库类型除书目数据库外,将朝着全文数据库、图像数据库、事实数据库、数值数据库共存的方向发展;其内容除科技外,将扩展到经济、管理、商业、法律和家庭等各个领域。通过情报检索系统,用户可以检索到所需的各种信息。

3.3 情报检索系统功能多样化

21 世纪的情报检索系统功能不再局限于印刷型文献信息的输入、处理、存贮、检索与输出,图像、电话、电视、缩微和机读等多种媒体的信息都将成为情报检索系统输入、处理、存贮、检索和输出的主要对象。其系统不仅具有灵活的数据组织能力和词表管理、自动切分与抽词、自动分类、自动标引、自动编制文摘、自动翻译等数据处理能力,而且具有情报检索、远程传真、电子邮件、高速打印、联机信息处理与服务等多种功能。

3.4 情报检索应用社会化

情报检索需求的社会化和情报检索技术的现代化将导致情报检索应用的社会化。首先,情报检索将类似通讯、煤气和水电一样成为社会的公共事业,用户不去图书馆,不求助情报人员,也能获得自己所需要的信息。其次,情报检索系统将从文献情报机构走向社会,渗透到社会生活的各个领域,成为社会机构中必不可少的组成部分。

3.5 情报检索系统建设整体化

随着情报检索系统发展战略研究的深入和宏观指导的强化,我国目前情报检索系统建设中存在的只注重开发试验、忽视实际应用、贪大求全、各自为政、工作相互重复、资源无法共享的现状将得到改善,情报检索系统建设将向着整体化的方向发展。一个以大型数据库为基础,兼具自动标引和翻译功能,集中型智能情报检索系统和无数以“光盘+微机”数据库为基础的小型分散式情报检索系

统同时并存,相互补充,共同发展的时代已经为期不远。

3.6 情报检索服务产业化

在我国社会主义市场经济体制不断完善和信息产业化程度日益提高的大趋势下,一方面,情报检索服务的商品属性逐渐为人们所认识,情报检索服务将走向市场,有偿服务将为用户所乐意接受;另一方面,数据库的生产、流通将会突破自给自足的小生产方式,而按照产业化的要求进行社会化大生产。数据库生产商、批发商、流通增值商、零售商和各种情报检索服务企业将陆续问世。情报检索服务产业将成为我国信息服务产业的支柱产业,在促进我国信息产业和国民经济的发展中将发挥重要作用。

参考文献

- 1 高崇谦.我国科技文献计算机标引研究的现状.情报学报,1989,8(1):67~73
- 2 曾民族.联机信息服务及其通讯网现状.现代图书情报技术,1994,(3):2~7
- 3 曾民族.我国计算机情报管理的技术进步和今后课题.

(上接第 8 页)

的研究,从而有力地推动了文献计量学、科学计量学的研究不断深入发展。通过文献计量学、科学技量学的课程教育,一方面培养了这些学科的专门人才,另一方面萌发了其他学科的人员对文献计量学、科学计量学的兴趣,从而投身到这些学科的研究队伍中来。这是我国文献计量学、科学计量学发展的一支重要力量。

参考文献

- 1 邱均平.我国文献计量学的研究和发展.情报学报,1987;6(6):466
- 2 赵红洲,蒋国华.科学计量学的历史和现状.科学学研究,1984,2(4)
- 3 汤世国.弗·威·兰卡斯特与普·阿·柯克媛.情报科

情报学报,1991,10(5):322~328

- 4 曾民族.中国 90 年代情报技术展望.图书情报技术,1990,(1):2~7
- 5 王能琴,陆长旭.多媒体技术的进展.现代图书情报技术,1994,(1)47~50
- 6 陈光祚.论数据库的二次开发与微数据库的建设.情报学刊,1993,12(5):365~368
- 7 高崇谦,张凤楼.中国计算机情报检索系统建设的现状和发展.情报学报,1989,8(6):460~465
- 8 周智佑.联机情报市场的分析与比较.情报科学,1988,9(3):61~69
- 9 郑登理.可擦重写相变光盘近期研究进展述评.现代图书情报技术,1993,(3)50~55

张帆 1966 年毕业于武大图书馆学专业,现为华中师大信息管理系主任。发文 25 篇,出版专著 1 种、译著 1 种。通讯地址:武昌珞珈路 100 号,邮编:430070。

姜策群 1982 年毕业于西安交大气体力工程专业,1986 年毕业于武大图书馆学专业,现为华中师大信息管理系副教授。发文 40 多篇,出版专著 5 种。通讯地址同上。

(来稿时间:1994-09-12。编发者:丘峰。)

学,1983,4(1):92

- 4 余惠芳译.日本图书馆情报大学本科专业课程介绍.情报科学,1983,4(6):69
- 5 Letter to the Editor Foundation of Information Science. J. Inf. Sci, 1982, 5(1):45
- 6 张保明.我国情报科学量化研究的进展.情报学报,1986,5(3~4):296
- 7 邱均平编著.文献计量学.北京:科学技术文献出版社,1988

邱均平 1969 年武大图书馆学系毕业,1981 年武大科技情报专业毕业。现为武大图书馆学院副教授。已发文近 50 篇,出版专著 7 种。主要著作有《科技文献学》、《文献计量学》等。通讯地址:武汉市,邮编:430072。

(来稿日期:1994-12-07。编发者:徐苇。)