

我国计算机情报检索的现状与未来的发展

王 兵

我们正处在一个情报爆炸的时代。

自本世纪五十年代以来,科学技术有了非常迅速的发展。新兴技术,如信息技术、生物工程、新型材料技术等相继出现。与之相对应的大量科技情报也如破竹之势。这些科技情报已经形成了一个巨大的冲击波,它不仅影响我们的物质生产,而且已渗透到了全社会的各个领域。如何从这些浩如烟海的科技文献中迅速、准确地找出符合人们需要的情报,这是计算机情报检索的最终目的。当前,在工业发达的国家里,计算机情报检索已成为人们日常生活中一个不可缺少的部分。大型国际联机情报检索系统,如DIALOG, ORBIT, ESA, MEDLINE等已做到全球情报资源共享,满足用户的实时性要求。在我国,我们也设立了国际联机情报检索终端,各大部委已基本上普及了脱机批处理式的情报检索,联机情报检索方兴未艾。但从整个现状来看,我国的计算机情报检索还很落后。根据这一情况,本文拟从三个方面来全面探讨我国计算机情报检索的现状与未来。

一、我国计算机情报检索的现状

自七十年代中期以来,计算机情报检索在我国有了迅速的发展,逐步从研究阶段转入实用阶段。据笔者调查和统计,除少数边远地区外,各省、市级的情报所都已在不同程度上开展了计算机情报检索。仅从在京的各部委情报所以及象中国科学院图书馆、北京图书馆等这样的大型公共图书馆来看,他们都已开展了计算机情报检索的研究工作,其中绝大部分都已开始向社会提供计算机情报检索服务,效果普遍反映较好。在引进国外数据库方面,我国已引进了33种国外文献磁带,订购了36种,其内容主要是化学、物理、机械、地质、海洋、生物、能源、农业等。每年约200多万条文献,累积文献总数已达2500多万条。目前,定题用户数已达到3000多个,课题总数为1万个。1980年,建工部在香港购买了一台终端,通过该终端与美国DIALOG和ORBIT两个大系统进行联机检索,其检索结果邮寄回国内。目前,它已为国内共累积检索了4千多个课题。为了进一步方便用户,我国在五机部和北京电报大楼又分别设立了DIALOG和ORBIT的电传终端,中国科学技术情报所也建立了ESA系统的终端。截至1984年11月止,ESA终端共检索了3512个课题,用户上机次数为1761次。各部委情报所的联机检索已达7000多个课题,广州新建终端也有近1000个课题,使用外汇达30万美元。这一切表明,计算机情报检索已在我国初具规模。

从1974年开始进行计算机情报检索的准备工作到现在,我国计算机情报检索从无到有,

从小到大。首先,我国情报界为计算机情报检索的前期工作做准备,掀起了一个编写主题词表的热潮,先后共编制和出版了若干综合性和专业性主题词表,到1984年底,正式的专业词表已达17部,其中有代表性的是综合性的汉语主题词表。最近几年,几百个单位在微型计算机上建立了试验性的或小型的情报检索系统,其中比较突出的是交通部情报所的微型计算机情报检索系统,他们将微型计算机与缩微、录相、光盘等高密度存贮技术相结合,发展存贮和检索一体化的系统,为在我国图书情报界应用微型计算机以及继续推广微型计算机的应用,使之成为大系统的后援,实现多机网络化,开辟了一条新途径。与此同时,上海交通大学等一批上海高校通过联合国拨款等多种经费途径,进行了计算机联网试验,以实现网络化检索,达到地区情报资源共享。在我国,数据库事业也有了比较大的发展,其中以中国科学院的数据库工程和高校学报文摘英文版磁带最为突出。事实数据库也在试验中,武汉大学中文系和计算机系合作建立了语言资料库,已存入了800万字,提供语言分析和词频研究。在数据库的建设上,从1978年开始,共召开了三次全国性的数据库学术会议。目前,除了积极利用各种渠道引进国外文献数据库外,已建成、拟建或正在建的数据库有31种,其中,有近20个中文数据库(不包括科学院系统)。数据库是计算机情报检索的基础,这些数据库的建立必将推动我国计算机情报检索的发展和普及。在学术理论研究方面,1981年以来,共召开了四次全国性的计算机情报检索学术交流会,参加的人数为501人,会上交流的论文达311篇,研究的范围从基础理论到具体应用的经验,内容丰富,涉及面广。1985年10月20—25日,在北京召开了有美国、日本、加拿大、欧洲航天局等国家和组织参加的国际计算机情报检索讨论会,这在我国还是第一次。与会的中外专家就当前计算机情报检索中的若干问题进行了探讨,对如何建立我国国家计算机情报检索网络这一专题进行了广泛的讨论。这对于我们今后更好地开展计算机情报检索的研究,有效地开展计算机情报服务都有着重要的现实意义。

二、我国计算机情报检索存在的问题

纵观我国计算机情报检索的现状,我们可以看到,我国的计算机情报检索确实有了很大的发展,成绩是可喜的。但我国计算机情报检索的整个水平还是不高的。通过自己在京期间两个多月的实地调查和阅读有关文献,对影响我国计算机情报检索的因素进行了归纳与总结。

A、从宏观横向面来看:

1) 缺少一个统一协调、统筹兼顾的全国性计算机情报检索的发展规划。

近几年,计算机情报检索发展迅速,在此基础上制定了六五计划。在这个计划中,虽然从总体上指出了今后工作开展的方向和重点,但如何具体地组织各省、市,地区级的计算机情报检索系统,怎样在国家科委的领导下协调各部委的情报工作,如何设计全国通讯网,如何建立与计算机情报检索系统相匹配的服务机构,以及怎样使计算机情报检索系统与经济信息系统相协调等等,这些问题都没有做出全盘的考虑和可行性研究。造成这种状态的主要原因是行政管理体制自身存在的缺陷。例如,各部委情报所属各部委领导,综合性科技情报所属国家科委领导,公共图书馆属文化部领导,而高校图书馆则属国家教委的直接领导。这种没有统一的领导机构来制定全盘工作和制定全面规划的局面,使得各部门只从本单位的利益出发力求“全”字,自成一体体系。这对于今后工作的开展,特别是全国计算机情报检索系统的建立,都带来了很大的困难。

2) 强调国外经验过多, 结合国情、研究实际情况不够。

充分借鉴国外的先进经验是必要的, 但国外的先进技术不一定都能适应我国的国情。目前, 在我国有一股盲目引“机”的倾向, 有些单位和部门, 只要有资金就想设法地从国外引进计算机和相应的软件, 甚至将全套设备引进, 忽视了天时、地利这两个因素。结合国情不够的另一个方面是对用户需求和用户需求的分布状态研究不够, 用户需求是开展情报检索服务的基础, 当前, 在已开展服务的计算机情报检索系统中, 一次性用户很多, 长期用户较少, 这说明用户对计算机情报检索服务带着一种试探性的目的, 还没有真正地认识到它的重要性。

3) 对计算机情报检索经济效益的研究不够。

计算机情报检索能节省科技人员查阅资料的大量时间, 准确、迅速地提供情报。从这一点上看, 它的社会效益是无可否认的。但过份地强调它的社会效益则不全面。前些年, 有关某次计算机情报检索的结果给生产带来很大效益的文章是很多的, 这样一来, 有许多单位就只看到它的社会效益, 而系统本身是否亏本, 建立系统是否值得等问题则考虑很少。因此, 除少数象中国建筑技术发展中心的单位开展服务略有剩余外, 其他绝大多数的部门在开展计算机情报检索服务时是亏本的。从国民经济发展的角度来看, 这种状况是不能长期存在下去的。当前, 我国计算机情报检索系统绝大多数都是由国家来补贴, 这样做对于发展计算机情报检索无疑是正确的和及时的, 但从长远利益来看, 从现在起, 我们就必须高度重视计算机情报检索系统自身的经济效益研究, 尽量减少国家的补贴, 提高计算机情报检索系统的经济效益。

4) 人员组成结构不合理, 单打一现象严重。

在我国计算机情报研究中, 人员分布状态很不适应现代情报研究的需要。主要研究人员是情报工作者, 而计算机、数学、系统工程等方面的专家则很少。应该看到, 在计算机情报检索中, 计算机人员、通讯人员等与情报工作者处于相同的地位。但我们目前的状态却是计算机人员的比重太小, 而且计算机人员只搞计算机, 很少考虑与情报人员的合作。至于数学、通讯、系统科学方面的专业人才则更少, 以致造成了目前这种单打一的现象, 加上现在存在的人才不流通的弊病, 因此, 更加促成了这一现象的发展。

5) 与计算机情报检索相匹配的原文提供系统不齐全, 原始文献提供困难。

当前, 计算机情报检索系统只能检索题录或文摘, 若进一步了解原文则无能为力, 虽然能向 DIALOG, ORBIT、ESA 等系统订购原文, 但这种方式昂贵, 不是长远之计。从北京各大部委的资料室、图书馆的文献资料收藏来看, 我国目前各种文献资料的收集还是比较齐全的, 如何将这些丰富的收藏合理地整理, 使之与计算机情报检索系统结合起来, 做到随时都能向用户提供原始文献, 这是当前计算机情报检索的一个薄弱环节。

6) 重点不突出, 造成展开的面大、难以合拢的局面。

目前, 重点工程强调不够, 没有侧重。造成这种局面的原因, 一方面是我们计划做得不细致, 不全面; 另一方面也是我们求胜心切, 实地调查和研究不够。摆在我们面前的任务很多, 数据库的建设、大型汉字软件、联机网络、机器翻译、检索语言等等, 这些工作都需要我们一个一个地解决。但先解决哪个, 怎样解决, 是否具备有足够的人力、物力和资金, 这些问题只有在做了全面细致的考虑后才能着手解决。

7) 分布不合理, 畸形发展。

当前, 大多数计算机情报检索系统都设在北京、上海等地, 边远地区几乎没有, 这样一来就容易造成大量人力、物力的浪费。造成这种状况的原因有客观的, 也有主观的。其客观

原因有两个方面。其一是我国人口众多,计算机普及率较低,有许多人不会使用或根本不知道计算机情报检索。其二是我国幅员辽阔,大量的现代化企业都设在大城市,加上边远地区交通不便,这样就给信息交流带来困难。若我们按点——点方式进行计算机情报检索,则计算机利用效率低,很不经济。其主观原因是我们对边远地区的情报资源的开发和利用不够重视。

B. 从微观纵向面来看:

1) 处理技术落后。

从总体上看,计算机的非数值处理比重很少。目前,计算机只能做定题服务,联机检索软件还处在引进和试验阶段,这阻碍了计算机情报检索的发展。

2) 系统组织比较混乱。

从经济学的观点来看,计算机情报检索系统是一个企业实体,没有一套合理的经营管理方法是不能生存的。但现有系统的各部分联系不明确,对系统的运行、费用等问题认识不清。

3) 系统人员管理水平较低。

当前的情报检索人员能以高度的准确性来认识机器的特性,但对人的特性却达不到同样好的程度。计算机情报检索系统的功能依赖于人这个控制器的特性。在系统人员组织管理方面,我国的水平还是比较低的,主要表现在:a. 规划低(制定目标、做出决策方面);b. 组织分工不明确(人员的分工、合作、协调、统筹等方面)。

4) 检索结果评价指标单一。

对于检索结果的评价,我国多采用查全率和查准率这两个指标。由于这两个指标都涉及到人这个因素,因此,在结果的评价上就存在着模糊性,仅从这两个指标来衡量系统运行的好坏是不充分的。目前,我国大多数计算机情报检索系统都使用这对指标来揭示本系统运行的效果,这就显得评价指标单一,不能全面地说明问题。但是,我们并不是因此而否认了查全率和查准率的使用价值。评价的目的,就是要通过用户对检索结果的评价,看到计算机情报检索系统的存在价值。其方法是多种多样的,最有效、最普通的方法就是经常向用户分发用户检索心理调查表。国外许多著名的大系统,如DIALOG和ORBIT,都是采用这个方法来调查用户的情报需求,用户对检索效果的看法,用户对系统的评价等问题。通过对调查结果的统计和分析,找出哪些影响系统发展的主要方面,提出系统的改进意见。

5) 理论与实际联系不够。

在我国计算机情报检索研究中,很大一部分研究都是探索性的研究,真正投入实际工作、特别是能导致实际工作有重大突破或进展的理论研究成果是很少的。这样一来,理论与具体实际工作相背离的现象就越来越严重,从而在我国情报界形成两股势力,一类是专门从事理论研究的人;另一类是专门进行实际工作的人。具体地讲,就是高等学校和各个实际的研究单位。

6) 检索费用没有一个统一的标准。

现在各个服务单位对用户的收费缺乏统一的标准,有些过高,有些则偏低,他们不是以系统运行状态的好坏来向用户收费,而是象征性地收取用户的检索费用。总体来看,就是对成本、费用以及它们之间的关系没有作认真的研究。

三、我国计算机情报检索今后的发展

以上我从横向、纵向共13个方面论述了目前我国计算机情报检索存在的问题,这些问题

就象一盘渗透到计算机情报检索各个环节的沙子，它不仅使得现有的计算机情报检索系统难以发挥正常的功能，而且它还阻碍了今后的发展。在这些多方面的问题中，如何区别根本性和枝节性问题是一个关键。只有将这两类不同性质的问题正确区分开以后，我们才能抓住牛鼻子，才能找出我国计算机情报检索今后发展的道路。下面我从两个方面、按重要程度的先后次序，阐述一下自己对我国计算机情报检索今后发展的看法。

(一) 计算机情报检索今后发展的根本性问题

计算机情报检索今后发展的根本性问题是今后工作的重点，这些根本性问题是纲，只有抓住了这个纲，其它问题才有可能迎刃而解。这条纲共包括三方面的内容：

(1) 制定出战略规划与政策

制定出计算机情报检索的战略规划是正确估量现实和有效地预测未来的重要保证。但制定这样的规划并不是一件容易的事情。首先，我们必须根据国家发展的需要，在规划中充分体现出国民经济规划的精神。具体讲就是要讲实效，不能浮夸，只顾高速度，而是要脚踏实地、扎扎实实地向前迈进。制定规划工作要有统一的领导机构，各个部门要从总体利益出发，克服本位主义，将规划中制定的各项任务落实到具体的工作中去。这就是说，各个部门要按规划中的规定，打破本部门“小而全”的体制，积极地向今后全国计算机情报网络的建设方向靠拢。在制定规划时，要做到统筹兼顾，兼顾长远利益与当前利益。在规划中，我们要正确处理长期战略与当前各个部门经营的矛盾。规划工作必须制度化、组织化。在制定规划的前期，要有组织地建立起规划调研小组、规划决策小组、规划管理小组以及规划信息小组。只有这样才能有的放矢，使制定规划的人有信息依据，从而保证规划工作的顺利进行。规划制定出来以后，若仅靠作决定、下命令来使各部门贯彻执行是行不通的。为此，必须对规划工作进行检查与控制，定期地对战略规划和政策的制订过程和实施结果进行评价。因此，必须建立起规划评价小组。这样，我们就可以在规划制订过程和结束以后，监督和保证规划的有效执行。在战略规划中，还必须体现出合理布局的精神，解决好集中与分散的矛盾，特别是要十分明确地指明国家、地区的各级计算机情报检索系统的任务以及它们之间的关系。按照“七五”计划中经济建设的战略布局的主要方针，根据我国经济布局客观上存在着东、中、西部三大地带的客观实际情况，为了改变目前大多数计算机情报检索系统集中于首都的状况，在规划中，我们要以东、中、西部为三个支撑点，层层展开，逐步地有计划地形成全国计算机情报检索网，这样，一方面能使全国性的计算机情报检索网有步骤地实现，另一方面也使得在建立全国性的计算机情报检索网的同时，使各级计算机情报检索系统充分发挥作用，以配合全国各地区的国民经济建设。最后，在规划中还要充分考虑计算机情报检索系统与其它系统、特别是与经济信息系统的兼容，这样，我们就可以做到一机多用，达到事半功倍的目的。

(2) 做好数据前处理工作，加快数据库建设

数据前处理工作是计算机情报检索的基础，数据库建设更是关键。根据我国的国情，我们在数据库的建设上是以引进国外数据库为主，还是以自建中文数据库为主，这是今后数据库建设的两个方向性问题。目前，国内对这个问题的看法也不统一。有的赞成引进，他们的主要理由是国外已有现成的数据库，而且已证明具有实际应用的价值，直接引进利用可以缩短全国计算机情报检索网络的建设周期和节省数据库的研究开支。也有一部分人赞成自建库，他们的主要理由是大多数科研人员不容易看懂外文，建立中文数据库是今后全面推广计算机情报检索的基础。

毫无疑问,数据库建设是当务之急。上面的两种看法各有其正确的一面,也有其不全面的一面。我认为,我们应从国情出发,有计划、有步骤地进行数据库建设,逐步实现具有我国特色的中文数据库,具体讲,就是有计划、有组织地、按先后次序建设数据库。即在全国计算机情报检索网络建设的同时,分期分批地适当引进国外数据库,这一方面能在引进的同时,充分发挥国外数据库的社会效益;另一方面也为我们今后与世界大型联机情报检索网络联网打下基础。然后在不断积累经验的情况下,根据当时的技术设备和客观条件,自建中文数据库。一旦中文数据库建成以后,我们就可以同有关国家交换数据库,进行国际合作。这样一来,我们就可以在不增加资金的情况下,在我国的计算机情报检索网中,既可以检索国外数据库,也可以检索中文数据库,使国产情报资源和国际情报资源为我所用。这样做有许多好处。首先,它能使有限的资金用在刀刃上。其次,我们能在国外数据库检索的实践中,逐步摸清数据库建设的经验。为什么要以这种方式来进行数据库的建设呢?这也是根据我国的国情所决定的。数据是数据库的基础,然而,我国目前的数据收集工作是做得不好的,标准化、规范化工作也不理想。这些工作都是数据库建设的前期准备工作,可想而知,这些工作不做好,又怎么能谈得上自建数据库?在中文数据库的建设上,我们也不能贪大求快,要从我国实际情报用户的需求出发,有选择地建立二次文献数据库。根据当前国民经济的建设情况,我们应尽快地建立起一、两个经济信息数据库,以适应国民经济建设的需要。

(3) 加强汉字情报检索软件的研制工作

计算机汉字情报检索不同于英文或其它文字的情报检索,这样就给我们的研究带来困难。同时,它也使得引进的情报检索软件难于对汉字进行检索。为了配合全国计算机情报检索网络的建设,我们必须尽快地研制出适应于汉字的情报检索软件。由于汉字是方块字,不仅输入难,检索也难。对于汉字情报检索软件来说,目前还存在着比较多的技术难点,其中有些是不可能短期内解决的。例如,汉字的切分问题。汉字的切分问题主要是影响汉字全文本数据库的检索。因此,我们一方面要组织人员研究汉字切分问题,另一方面也要注意汉字书目数据库和数值数据库的检索软件的研制。在汉字情报检索软件的研制过程中,我们要特别注意汇集各方面的人才,特别是语言专家和软件专家,克服目前一边倒的趋势,即只有情报人员来完成这项工作。从当前的研究状况来看,大部分的语言和计算机专家将注意力集中在汉字输入方案的研究和比较上,而对于汉字的检索软件则考虑较少,这就使得这项工作进展缓慢。我认为,我们应高度重视这个问题,要由国家有关部门统一调配人员,形成一个由语言专家、软件专家和情报人员共同参加的三位一体的研究小组,这样协同攻关,早日实现具有中国特色的汉字情报检索软件。

(二) 进一步完善我国计算机情报检索应注意的方面

以上三个方面的问题是我国计算机情报检索今后发展应高度重视的根本性问题。下面将要论述的是一些非根本性的问题,这些非根本性的问题并不是说不重要,它是相对于根本性问题而言的,在根本性问题解决以后,这些非根本性问题就变得突出了。因此,我们在集中力量解决这三个根本性问题的同时,还必须考虑这些非根本性问题。这样,我们就能不断完善我国的计算机情报检索事业,使之真正成为为四化建设服务的有力工具。

(1) 加强计算机情报检索系统经济效益的研究

经济效益的好坏是确定系统生成、发展、壮大的必要条件。在上述问题解决之后,经济效益的研究就变得至关重要。这里面要着重进行系统质量和费用管理方面的研究,彻底改变我国计算机情报检索亏本的局面。

(2) 加强计算机情报检索系统的组织与管理

计算机情报检索系统一般是由计算机系统、系统人员(标引人员、构造检索策略人员、维护人员等)和用户等部分组成。因此,加强系统组织与管理的关键,首先是要组织、协调好系统人员;要使计算机系统在整个系统人员的管理下,输出用户满意的情报,其次是尽可能地按现代企业管理的方法应用到计算机情报检索系统的组织与管理上。

(3) 积极做好各种人员的培训工作

在整个计算机情报检索中,人始终起主导作用,积极做好各种人员的培训工作是维护系统正常运行的关键。在今后计算机情报检索服务中,不仅要努力做好用户的培训工作,而且还要做好系统人员的培养工作。国外是十分注意这方面的投资的。因此,在未来计算机情报检索的发展过程中,要抽出一定比例的资金来开展这方面的工作。

(4) 建立起有效的信息反馈系统和辅助系统

所谓有效地开展计算机情报检索服务,就是要随时掌握计算机情报检索系统运行的情况。这就要求我们建立一套行之有效的信息反馈系统。这种反馈系统能及时地反映出用户的各种要求。要做到这一点,我们必须克服检索结果评价指标单一的现象,设立用户咨询小组和检索结果评判小组,及时地对用户的检索结果进行评价,做到眼观六路、耳听八方,以达到最优服务。与计算机情报检索系统相匹配的辅助系统是有效开展服务的重要保证之一,从今后的发展来看,我们应努力做好原始文献的提供工作,这就需要我们从现在起积极地筹备全国馆藏目录的编制,为今后开展计算机情报检索服务提供方便。

(5) 加强新技术的开发和理论研究工作,逐步实现全国联机情报检索网

当前,以信息科学、微电子技术为核心内容的一场新技术革命正在世界各国蓬勃兴起,计算机情报检索也不例外,传统的检索方式正在逐渐被淘汰,新型的检索模式,如非布尔式检索、概率检索、模糊检索等,已在试验过程中。所以,我们要看准这个机会,借这次新技术革命的浪潮,把计算机情报检索的理论研究和新技术开发搞上去。对新兴课题,如非布尔检索、相关性检索、自动语言处理、专家系统等,要集中人力、物力重点攻关。在进行理论探讨的同时,要逐步地实现全国计算机情报检索的网络化。网络化的关键是通讯技术,在进行通讯技术的研究时,要特别注意开发光导纤维通讯技术。在我国,目前拥有10多万台微型计算机,这些微型计算机是今后全国计算机联机情报检索的物质基础。因此,从现在起,我们就必须有计划地进行微型计算机与大型计算机的联网试验,逐步统一微型计算机的机型。国外曾经有人预言过,微型计算机、光导通讯是新时期计算机情报检索的支柱。现在看来,我们有数目可观的微型计算机和大型计算机,只要我们计划协调工作做得好,实现全国计算机联机情报检索的日期是不远的。

参考文献:

- (1) F·W·兰卡斯托著,陈光祚译:《情报检索系统:特性、试验与评价》。
- (2) 曾民族:“建设数据库是信息系统现代化重要任务之一”,《情报学报》1985年第4卷第1期。
- (3) 曾民族:“我国计算机情报检索的发展和研究课题”,《情报学报》1985年第4卷第2期。
- (4) 王兵:“情报检索系统评价”,武汉大学硕士研究生学位论文,1985年。
- (5) H·S·希普斯著,张承庆、顾慧芳译:《计算机情报检索导论》,知识出版社1984年。
- (6) 李鸿梁等著:《国际联机情报检索》。