

联机情报检索的现状与展望

陈 光 祚

所谓联机检索,就是用户通过终端设备同检索系统相联结而进行文献与数据的检索。这种联结是借助于通讯路线实现的。尽管用户同检索系统的中央计算机之间相隔千山万水,却可以如同近在咫尺一样地进行“对话”。用户通过终端的键盘,把自己的检索提问输入计算机,而计算机则立即响应,及时地检索存贮在数据库(即大型资料档)中的有关情报,把检索结果显示在终端的屏幕上,或用电传机打印出来。用户可以随时修改自己的检索范围和角度,请求计算机协助自己挑选最合适的检索词,浏览和筛选计算机检索的初步结果,从而实现“即问即答”的检索过程。

联机检索的过程,大致如下:

首先,用户拿起电话,叫通接至检索系统的通讯线路。

接着,在终端的屏幕上显示出来自检索系统计算机的信息:你好!这里是××××检索系统。现在的时间是×时×分。请把你的口令(即终端暗号)打进来。

用户打进了口令。计算机确认这个终端有权对其进行检索后,就向用户询问:你是一个有经验的用户吗?如果是,请打进一个“Y”;如果否,则打进个“N”。

如果是一个无使用终端经验的新用户,计算机将首先告诉他如何使用的方法。如果是一个老用户,打进一个“Y”之后,计算机就向该用户显示自己有哪些可供检索的资料档,并问用户需检索什么资料档?

用户根据自己检索课题,选择自己所要的数据库(资料档),并通过键盘打进资料档的编号、名称。

然后,计算机就询问用户的检索语句是什么。用户将检索提问通过标准的检索词及其正确的组配关系组成检索语句。用键盘输入。这样,计算机就根据检索语句中的检索词及其逻辑关系进行查寻工作,把资料档中符合该检索语句的文献或数据查找出来。并告诉用户共有多少篇文献,题目是什么?篇名、著者、出处是什么?必要时并可输出摘要全文。一般来说,用户首先要求计算机输出文献篇名,以便从题目进行核对或筛选。然后要求计算机把选定的文献的其它著录事项或摘要输出。

检索结束后,计算机自动统计用机时间和费用,并最后输出“再见”字样。

联机检索是继批式检索之后出现的,联机检索服务部门从产生到现在已有十三、四年的历史了。在这段时间里,联机检索的技术逐步得到改进,联机检索系统逐步增加,而成为当代图书、情报工作中的一个重要方面。

与批式检索比较起来,联机检索的优越性是很明显的:

第一,联机检索可以帮助用户得到更准确的情报资料,而可避免大量的无关情报。当用户不知用什么适当的检索词表达自己的检索提问时,系统可以向用户展示一系列有关的检索词或范畴表,供用户参考和选择,启发用户的检索思路,逐步修正或明确自己的检索语句。这样保证了正确的检索策略的形成和良好的检索结果的取得。

第二,联机检索取得情报的速度很快。一般不十分复杂的检索提问,从第一个检索词的输入,到检索结果的最后一个字的输出,往往不到一分钟。如果用手工检索,则需几小时甚至几天的时间。即使是批式检索,用户也不能“立等即得”。

第三,联机检索可以打破地理上的隔阂,使远离计算机中心的用户可以进行情报检索。用户即使是在一个穷乡僻壤的小图书馆里,也可利用终端查到如同在大图书馆找到的同样丰富的文献资料或科学数据。从这个意义上来说,大量的文献资料就在用户自己的手指头上。他只要按动键盘,就可以获得该检索系统所拥有的全部数据库中的有关资料。

第四,联机检索的费用不算太高。费用同用机时间有关,大约是一分钟一美元。同时检索费也同终端至系统的中央计算机的距离有关,距离越远,所需的通讯费用就越高。但总的来说,用不到一分钟的时间,从几百万条记录中检索出自己需要的有关资料,并且能及时地打印出来,从而节约了大量的人力与时间。从这个角度看来,联机检索的费用不算太高。

早期的联机情报检索系统是由私人或商业性的机构研制的。例如今天的美国系统发展公司的 ORBIT 系统就是在六十年代初期由该公司的科学家研制的。其后由美国国防部组织了全国规模的试验。经修改后,被美国国立医学图书馆所采用,而成为 MEDLINE 系统。今天,至少有十五个国际规模的或具有同等潜力的联机检索系统,其中有十二个建在美国。大部分联机检索系统是商业公司性质。有名的有:系统发展公司、洛克希德公司、俄亥俄大学图书馆中心、纽约时报、米德公司、BATTELLE 纪念研究院、情报学与目录学检索服务处等等。除商业性联机检索系统之外,欧洲航天局联机情报检索系统、英国国家图书馆自动化情报服务系统、美国国立医学图书馆联机检索系统等等,都是世界上较重要的联机检索系统。大多数欧洲的联机系统是由政府兴办的。

联机检索系统的关键技术是分时计算机、高速磁盘存贮器以及电讯设施。分时计算机可以同时为许多用户进行检索。近年来,各系统的中央计算机几经升级,内存与外存容量扩大几倍、十几倍甚至几十倍,终端数目每年几乎成倍增加。高速磁盘可以存贮几兆字符,计算机在磁盘中寻找其中任何一个字,大约只要 20 毫秒(即五十分之一秒)。许多公用通讯网络例如 Tymnet 和 Telenet 的建成,大大扩展了联机检索的服务范围。

由于多数联机检索服务是商业性的,彼此的竞争很剧烈。各方都力图改进功能,提高服务质量,以争取用户,压倒对方,从而导致联机检索新的功能不断出现。近年来,这些改进主要是:

延长服务时间。目前美国主要的联机检索服务每天至少开放十五小时,以适应世界各地由于时差而引起的在检索时间上的差异;

保存查找公式、提供联机定题服务;

左侧截词检索的能力；

文件排序的能力；

即日完成脱机打印，即利用高速线路和晚上低峰时间，为用户打印篇幅较大的检索结果，以降低用户的费用。

当然，上述新特点，并不是每个联机系统都同时具备的，有的系统具备这一、两个特点，有的具备另一两个特点。

目前，各联机检索系统所能提供检索的数据库（资料档）的数目，据有人在1977年作出的估计如下表：

	数据库数量	数据库的百分比	记录的数量	记录的百分比
美 国	160	58	46.3M	89
美国以外	117	42	5.7M	11
	277	100	52.0M	100

这就是说，他估计世界联机数据库共约277个，其中的58%，即160个是美国的。美国之外的数据库为117个，占总数的42%。联机数据库中的记录总额达52M（即5200万）条，其中美国是46.3M，占89%，美国以外的是5.7M，占11%。

科技方面的文献数据库主要有：《化学文摘选要》（Chemical Abstracts Condensates），《生物学科学情报预告》（BIOSIS Previews），《计算机化的工程索引》（COMPENDEX），以及英国物理、电工和计算机及控制情报服务（INSPEC）等等。

社会科学方面的文献数据库主要有：《教育研究情报》（ERIC），《心理学文摘》（Psychological Abstracts），以及《社会科学引文索引》（Social Science Citation Index）等等。

商业方面的文献数据库主要有：《会计师索引》（Accountants Index），《化学工业摘记》（Chemical Industry Notes），《药业新闻索引》（Pharmaceutic News Index），以及《世界专利索引》（World Patent Index）等。

多学科的文献数据库主要有：《美国国家技术情报服务》（NTIS），《Smithsonian科学情报交流数据库》（The data base of the Smithsonian Science Information Exchange），以及《美国国会情报服务索引》（Congressional Information Service Index）等。

近年来，从一个终端上可以利用的数据库的范围正在扩大。五、六年前，一般的商业性联机检索服务向用户提供二、三个数据库。到1975年，所提供的联机数据库的数目已增至15—20个。现在，主要的联机检索系统已向用户提供30—50个数据库，甚至更多。用户可以同几个检索系统订立合同，在一个终端上可存取近一百个数据库。现在摆在用户面前的问题，不再是一个获得存取的问题，而是如何选择最适合自己的系统和数据库的问题。

除了文献数据库或者说书目数据库之外，科学数值数据库也增加很快，并供联机检索。例如关于各种科技参数、化学品结构式、产品规格、制造厂商名录以及有关图象等数据库，可供用户检索具体的数据或图象。欧洲航天文献中心编制的“电子部件数据库”、“地球网络图象数据库”，帝国化工公司有机部的“精密化学品数据库”，法国的“热数据

库”等等，都是例子。

近年来，另一个变化是，用户使用联机检索的费用不断下降。尽管检索系统的数据库本身在持续地增加，需要更多的磁盘存储空间和花费更多的计算机处理时间，但由于计算机、磁盘存储器、电讯设施和服务工作的成本逐年下降，因此每次检索的平均费用正在稳定地下降。

近年来的情况还说明，联机检索效能的提高，一方面固然同计算机、磁盘驱动器或电讯设施的改进有关，但更主要的是联机检索系统的人员注意了解用户的需要和存在问题，并把这些需要和这些问题的解决方法，不断地变成程序的新功能，日益提高了软件的水平。例如字符串检索、词典显示、查找公式的保存等等功能的出现，使联机检索应用软件日趋完善。

联机检索的未来发展趋向是什么？例如五年、十年后，它究竟是怎样的？这个问题有一些推测与预言。

一种回答是今后将有更多的数据库，更多的系统，更长的服务时间，更快的打印输出服务，更低的费用等等。当然，这些是不成问题的。

今后，计算机的速度会变得更高。它将消除各系统之间由于设计和程序的好坏所造成的在响应时间方面的差别。同时我们也将看到各个系统之间的容量将逐步拉平。当出现这种情况时，改进的重点将越来越放在使系统更便于使用和成本更加下降方面。

使联机检索系统更便于使用的一个发展趋势是研制联机上下文关键词索引(On-line KWIC Index)。这种索引正在试验，不久将公开使用。美国系统发展公司在其 ORBIT 联机检索系统上将 NTIS 和 COMPENDEX 两种数据库进行编辑统一的上下文关键词索引的试验。它把每篇文献题目按关键词进行轮排，如同目前所有的上下文关键词索引一样。联机上下文关键词索引的新奇之处，是它可以包含任何数目的数据库。一种专门的程序将许多不同的数据库合并成单一的上下文关键词索引。用户可以将这些索引项目脱机打印出来，成为单独的书目。各个用户按不同的关键词进行检索，可以分别获得适合自己需要的、取自各个不同数据库的书目情报。因此，联机上下文关键词索引在本质上是一种新的工具，一种新的输出形式。用户对它的反应是良好的。

使联机检索系统更便于利用的另一改进，是今后几年内使联机检索系统成为因人而异的系统。也就是说，当住在美国、英国和加拿大西部的说英语的用户来检索时，系统用英语和用户进行对话。而当住在西班牙语地区的说西班牙语的用户检索时，系统就用西班牙语和用户对话。同样，对说德语的、意大利语的用户，系统就分别用德语、意大利语与之对话。当然，检索出来的文献题录或摘要仍然是原来的语种。因此用户需要懂得英语，才能了解从英美联机系统检索出来的结果。但与计算机对话，则可用他最熟悉的本国语言。这种因人而异的系统，对于检索开始与结束的时间，也以用户所在地的当地时间来记录。例如对同时来自洛杉矶和马德里的用户，系统记录的时间是不同的，例如前者为太平洋时间一时半，而后者为西班牙当地时间九时半。

使联机检索系统更便于使用的改进，还在于显示终端将变得更轻便、更便宜。但是轻便式的终端需要更小的屏幕。用户将向系统要求 40 字符或 47 字符、63 字符长度的输出，以取代目前 72 字符的标准长度。

(下转第 67 页)

献文帝时,大概一向就这样。

③⑧ 《宋书》卷 98《氏传》就提到晋宋之间有“四方流民”进入仇池。《北史》卷 96《氏传》也说“先时,四方流入以仇池丰实,多往依附。”仇池地区虽是氏羌聚居地,也有汉人错居是无疑的。又《水经注》卷 17《渭水篇》有长蛇戍,称“魏(文成帝)和平三年(公元 462 年),徙诸流民以遏陇寇。”《魏书》卷 30《陆真传》亦载,高宗文成帝时,“是时,初置长蛇镇,真率众……卒城长蛇而还。”则陇砦东西,本有迁徙流民或俘虏置镇立戍之共同特点。

③⑨ 参考吴廷璠《元魏方镇年表》(《二十五史补编》第四册)。

④⑩ 《魏书》卷 51《吕罗汉传》。

④⑪ 《魏书》卷 106《地形志》作“渠州”,误。

④⑫ 均见《水经注》卷 20《漾水篇》,参看《中国历史地图集》第四册,48—49 图。

④⑬ 杨氏自分宗党由来已久,杨难当和杨保宗之间的矛盾就很明显。杨文德依靠刘宋于元嘉 27 年(魏真君 11 年,公元 450 年)自汉中西上,《宋书》卷 98《氏传》说:“文德宗人杨高率阴平、平武诸氏据唐鲁桥以拒文德。文德水陆俱攻,大破之。”显然杨高是站在北魏一边的。

④⑭ 《魏书》卷 66《地形志》下“南秦州”条:“正始初置,治骆谷城。”

④⑮ 《魏书》卷 19 上《元丽传》。

(上接第 52 页)

当然,还可能有许多简化联机检索使用方法的变革。例如在联机条件下帮助用户从数百个数据库中进行选择最合适的数据库的方法。将出现把用户输入的词转换成数据库中有控制的词表中的规范词的机构。还可能出现把一种系统的指令语言中的词转换成另一种用户所不熟悉的系统所要求的词的机构。最后,有可能期望消除目前存在的数据库的生产者和联机检索服务供应商之间的界限,使之成为整体化的联机检索服务工业,等等。

联机检索系统的另一个发展趋势,是它也担负起联机编目的职能。即这个系统既可用来进行情报的检索,也可用来进行图书的编目。英国国家图书馆自动化情报服务——BLAISE 就是这样的一个联机系统。它除了提供检索的功能之外,各个图书馆也可以利用这个系统所拥有的美国国会图书馆的马尔克和英国马尔克的数据库,核对本馆入藏的新书,联机打印出新书的目录卡片,或向该系统订购计算机缩微输出胶片(COM)目录。这种兼备检索与编目两大功能的联机系统,看来是有广阔前途的。

联机检索系统逐步走向网络化与国际化,看来也是一个确定的发展趋势。这在西欧表现得特别明显。在美国,由于各个大型联机检索系统多数是商业性公司经营的,彼此竞争,互相抵销了不少人力物力。而在西欧,目前正在走情报工作集中化与网络化的道路,有计划地把现有的各联机系统组成网络。例如英国目前正在把国内已有的四大家联机检索系统——不列颠图书馆自动化情报服务系统(BLAISE),英国电气工程师学会的INSPEC系统,德温特专利情报检索系统以及化学情报服务部系统——组成一个叫INFOLINE的网络;而INFOLINE这个全英国的网络,又将纳入全欧经济共同体联机情报检索的大网络——EURONET。这种由单个的联机系统向包含多系统的大网络过渡,有利于克服重复浪费,使联机情报检索的建设达到计划化、协调化和科学化。