

论计算机情报检索工业

陈光祚

本文分析了作为信息工业一部分的计算机情报检索工业的形成及其结构,着重探讨了情报数据库生产和联机情报检索服务业的发展、内部竞争机制以及当前的格局。文章最后联系我国实际,提出我国建立发展这种工业的一些有关问题。

一、计算机情报检索工业已形成相当规模

计算机情报检索是在手工检索的基础上,采用计算机等现代化信息存贮与检索手段,使情报检索变成高速、准确与自动的过程。各种情报信息一旦成为计算机中的数据库,在软件的支持下,就能被检索、排序、组合和得到控制,真正实现了可操纵性。

作为一门学科的计算机情报检索,开始出现于50年代后期和60年代初期。它是计算机科学技术、电讯科学技术在文献情报领域中应用而导致产生的学科。大致来说,60年代是脱机批处理检索的技术时代,70年代是联机检索的技术时代,80年代则是步入网络化检索技术的时代。

作为一门学科的计算机情报检索的发展,直接导致图书馆和情报资料单位的情报服务新项目——计算机情报检索服务的建立。这种新型服务,改变了图书情报部门那种各自为战的手工式的传统情报服务,极大地提高了检索的效率,为情报用户开拓更为广泛有效的情报获取模式,并改变了他们的情报搜集行为,激发他们表现出更高的情报意识与情报需求。图书情报部门的这种新型检索服务项目,使计算机情报检索成为社会化的一项科学交流活动。

随着计算机情报检索服务的扩大与深化,情报用户对这种服务需求的增长,导致社会经济领域产生一种新的工业部门——计算机情报检索工业,或称联机情报检索工业。它包括:情报数据库的生产业、联机情报检索系统经营业、为联机情报检索提供必需的电讯服务的电讯业、协助情报用户从计算机情报检索系统进行检索的情报经纪人和中间人的代检服务业,以及计算机情报检索工业的顾客——情报最终用户群。在以上几个部分中,存在着情报信息市场的各种商品关系。这个工业也象社会经济领域的其他部门一样,充满投资、竞争、兼并和对利润的追求,同时也对社会的经济活动和科学研究起着信息保证的巨大作用。这种学科——服务——工业之间的关系,体现了高新技术迅速转化为社会生产力的过程。反过来,作为一种工业部门的计算机情报工业的发展,不仅为学科的发展提供实践的环境,而且从经济上有力地支持了计算机情报检索学科研究的发展。从而促进了学科——服务——工业三者之间的良性循环。

当前,计算机情报检索工业已形成相当规模。据 Cuadra 的《联机数据库名录》的统计:

种类 \ 年份	1979/80	1980/81	1981/82	1982/83	1983/84	1984/85	1986	1987	1988	1989	
数据库生产者	221	340	512	718	927	1189	1379	1568	1685		
联机情报数据库	400	600	965	1350	1879	2453	2901	3369	3699	4465	
联机检索系统	59	93	170	213	272	362	454	528	555	644	
联机关							35		59		

(注)单位:个

计算机联机检索用户及其检索量逐年都在增长,据英国 Aslib(英国情报学家协会)1988年的统计,工商业界在1982年有39.7%使用联机系统,而至1987年,这个数字已上升到78%;大学等,1982年为25.2%,1987年为60%;政府部门,1982年为19.7%,1987年为80%;公共图书馆,1982年为8.6%,1987年为63%;专业协会与学会,1982年为6.8%,1987年为55%。

联机情报服务的总营业额,美国约为70亿美元,欧洲为10—15亿美元,日本为5亿美元,其它国家为5千万美元。营业额的主要部分来自对财政金融和经济数据库的服务。

数据库的生产,美国占56%,欧洲占26%,日本占15.6%,其它国家占1.4%。联机情报工业每年的增长率为25%。

二、情报数据库生产业

情报数据库是整个计算机情报检索系统的基础。情报数据库生产业溯源于60年代。美国化学会下属《化学文摘》社在60年代初为改造印刷版《化学文摘》生产程序,而实行计算机辅助排版和关键词索引的自动编制,先将文摘输入计算机,成为《化学文摘》的电子文本,即机读书目数据库的雏形。美国国会图书馆研制开发了机读目录(MARC),解决了书目数据库的结构模式。从此,书目数据库成为情报数据库中首先出现的一种数据库类型,而学术团体和政府图书情报服务部门成为数据库的第一批生产者。学术团体成为首批的数据库生产者,并不是偶然的,一是他们有长期的二次文献工作的经验和基础;二是他们身处学术环境,对科学的进展和新技术的应用比较敏感;三是具备相应的人才。学术团体所生产的数据库,其特点是:数据库的选题范围以面向学科为主,即根据学科体系的分野作为数据库收录材料的依据;数据库的类型以书目数据库为主,在服务对象上以学术研究界为主;在经营目标上是非赢利性的。

政府机构以自己的职能范围为目标,借助于政府的力量和政策的保护,以及自身掌握大量的信息等优势,在数据库生产者中占有举足轻重的地位。象美国的商业部、农业部、运输部、能源部、环保局、地质局、宇航局等等,都是活跃的数据库生产者。他们所生产的数据库特点在于:数据库的主题范围以面向任务为主;数据库类型趋向多样化,即除书目数据库之外,还生产全文型、名录型等数据库;服务对象也从学术界扩大到管理部门和社会公众。

私营的数据库生产商异军突起。他们针对迅速扩大的情报市场,对数据库生产业进行投资,以赢利为目标而进行经营。与学术团体和政府部门比较,其劳动生产率较高。他们生产的数据库,特点在于:在数据库的主题范围上以面向市场为主;在数据库的类型上以源情报数据库(数据数据库、事实数据库、全文数据库)为主;在服务对象上以从事财经商业活动人员及一般公众为主。象 Data Coruier 公司、Predicasts 公司、Harfax 公司等都是极为活跃的数据库生产商。这种私营数据库生产商的出现,虽较学术团体和政府部门为晚,但后来居上,他们的生产能力,目前在数据库产品的比例上已占多数。他们的出现与发展,加速了数据库商品化的过程,改变了早期数据库编制依赖政府资助的状况,使其具备了自立的能力,使得作为信息产业一部分的数据库工业形成了规模,促使数据库市场、数据库贸易的发展。

近年来数据库生产者阵营的一个显著变化是新闻出版机构的涉足,他们的重要性正与日俱增。例如 Grilier

电子出版公司、Elsevier 科学出版社、牛津大学出版社、ORYX 出版社、NILS 出版公司、Wiley 电子出版公司、R. R. Bowler 公司、Thomas 出版公司、华盛顿邮报公司以及法国通讯社、国际联合通讯社等等都是活跃的数据库生产者。他们生产的数据库,尤其以新闻全文数据库、电讯稿全文数据库、杂志全文数据库以及工具书全文数据库等令人瞩目。新闻出版机构近年来致力于把它们的印刷出版物转化成电子出版物。

数据库生产的发展动向,越来越受信息市场力量的支配。那些需求量大、最能获得利润的数据库被置于优先发展的地位。在数据库的主题范围上,财经商业数据库的发展速度遥遥领先于科技数据库。经济新闻、公司名录、产品信息、金融信息、经营管理信息等等数据库日趋繁荣。在数据库的类型上,源数据库的数量已超过书目数据库。这是因为源数据库提供的是“纯情报”,它所提供情报的深度要比书目数据库大得多。在生活节奏加快的时代,源情报的直接可用性更适合情报用户的需求。在服务对象上,源数据库既能为学术研究工作服务,又能为广大公众所接受。因而有更宽的用户面。从商业竞争的角度来看,在书目数据库领域长期以来已形成象 CA—SEARCH(化学文摘)、BIOSIS(生物学文摘)、INSPEC(科学文摘)等权威性高和累积期长的数据库,很难使新的数据库生产者与之抗衡。然而源数据库领域则是一个广阔天地,可生产的数据库主题与类型繁多,正是商业投资的良好场所。所有上述各点都导致源数据库的迅速增长。

三、联机情报检索系统经营业

数据库生产者生产的数据库,一般以磁带形式发行或出租,不能直接被情报的最终用户所利用。这就产生了计算机情报检索工业中的第二个环节——联机情报检索系统的经营者。这种经营者拥有强大的计算机设备和情报检索软件,但一般自己并不生产数据库,而是向数据库生产者租用数据库,在自己的大型计算机上装上这些数据库,并与电讯公司订立协议,以取得电讯设施的使用权,从而向世界各地的终端用户提供有偿的检索服务。如果说数据库生产业是计算机情报检索工业中的基础部分,则联机情报检索系统是这种工业中的核心部分。

联机情报检索系统的经营,与数据库的生产比较起来,需要更多的资金与设备的投入。因此,联机情报检索系统的经营者必须有雄厚资金的公司或国家机构甚至国际机构来支持。

1966年,洛克希德导弹与空间公司创立了它的 DIACOG 系统,1972年,它以少数几个数据库开始商业性经营,向社会提供联机检索服务,至90年代初,它已装有近400个数据库,拥有8万多个口令字(即用户数)。DIACOG 的资产,1972年约50万美元,1981年增至7000万美元,1988年即增至3亿5千多万美元。同年 DIACOG 系统出售给 Knight—Ridder 公司,其后又售给 Maxwell 公司,但仍保留 DIALOG 的名称。DIALOG 虽装有许多重要的书目数据库,但现在它的一大半数据库是源数据库。它以商业和财经界作为主要的市场对象。例如,它建立了“DIALOG 商业连接”服务,这是一种为商业界用户服务的项目,并建立了通向证券交易所的计算机网关。近年来,DIALOG 的服务趋向多样化,如1982年起建立了“知识索引”服务,这是面向最终用户的晚间检索服务,收费低廉,检索指令简单易用,能提供数十个常用数据库的检索。此外,还建立联机订购(即文献提供)服务、电子邮件服务、DIALOG 医学连接服务、以及一批光盘数据库系统产品(称为 DIALOG OnDisc)。

系统发展公司(SDC)早在1965年就开始研制联机情报检索系统,并建成 ORBIT(联机分时书目情报检索系统)。该系统主要提供科学技术领域的数据库检索服务。美国国家医学图书馆经营的 MEDLIN 系统(联机医学分析与检索系统),既是首批联机数据库之一,同时也是世界上使用量最多的联机系统。BRS 联机情报检索系统于1977年以10个数据库开始投入营运。今天已提供100多个数据库,主要是书目数据库,但也有若干全文数据库。1988年12月,它被 Maxwell 通讯公司所收购。

在欧洲,依靠国际机构或国家情报中心建立的联机检索系统也纷纷出现。例如,由12个国家组建的欧洲航天局的 ESA—IRS 系统。目前提供100多个数据库的检索服务。

Data—star 系统建于1980年,原先为瑞士政府所属,1988年为 Radio—Suisse 公司所收购。该系统有100多个数据库,侧重于物理学、生命科学、社会科学,但也包括财经方面的数据库和全文数据库。1986年,它收购了欧洲石油化学协会的“欧洲贸易统计”数据库,从而也成为数据库的生产者。

法国政府1979年建立 Telesystemes—Questel 系统,装有70多个数据库,主要是科技领域的,但也有一些是

商业领域的。

值得注意的是 80 年代由美国化学文摘社、日本科技情报中心及德国 FIZ·卡尔斯鲁厄三家数据库生产者合作建立的 STN 系统(国际科技情报检索系统)。这个联机系统有三个中心,分别设在美国的哥伦布、日本的东京和德国的卡尔斯鲁厄,它们之间通过卫星进行通讯连接。这个系统提供的基于数据库,其中不少是这三家自身生产的。数据库生产者经营联机检索系统,有其“产销一条龙”的优势。

除上述联机情报检索系统之外,还有大量的系统存在。有的以提供源数据库检索服务为特色,有的以提供专门领域数据库检索服务为特色。

米德数据中心(Mead Data Centra)是重要的源数据库检索系统的经营者。目前提供两个系列的数据库及其检索服务,一是 LEXIS 系列(法律方面),于 1974 年开始提供商业性的检索服务。二是 NEXIS 系列(非法律方面,主要是新闻领域的),它于 1983 年开始营运。

一些专营财经数据库联机检索服务的公司及其系统应运而生。在这个领域内,原先从事计算服务的分时公司率先经营财经数据库的联机检索服务。他们的系统适合提供经济时列的数据格式及其运算的要求,有的软件能处理经济活动的数学模型。这是它们的软件优势。

新闻出版机构涉足联机情报检索系统的经营,其方式是收购现有的情报检索系统,对其进行改名和重组,从而成为联机情报检索系统经营者阵营中的重要方面军,这是 80 年代以来联机情报工业中的一个带趋向性的变化。

路透社(Reuters)是 19 世纪建立的著名新闻社,从 1964 年起也参与财政数据的计算机传输工作。它在 1987 年收购了 Finsbury 数据服务系统,1986 年以 4800 万英镑收购了 I. P. Sharp 公司的系统,主要提供商业信息服务。

培加蒙出版社(Pergamon Press)是著名的出版公司。它于 1980 年收购了欧洲的 Infoline 情报检索有限公司,将其改名为 Pergamon Infoline 公司。在 1986 年,培加蒙出版社又收购了系统发展公司的 ORBIT 检索服务系统。将其改名为 Pergamin.Orbit Infoline。

英国电讯公司(British Telecom)也参与联机情报检索工业。英国电讯公司还把它所属的几个信息服务——如“商业指导”、电子邮政服务(Telecom Gold)、电子图文系统(Prestel)结合起来,置于该公司的 Dialcom 系统的名下。

联机情报检索系统日益增多,它们之间的竞争也日趋激烈。竞争的方式是多样的。有的以所提供的数据库众多取胜。例如 Dialog 系统以信息的“超级市场”为标榜,吸引大量的情报用户。许多联机情报检索系统也在仿效,他们装载的数据库数目正在与日俱增。系统的信息资源的丰富程度确实是用户选择使用检索系统的出发点之一。

有的以提供独占性的数据库为号召,也就是说,这些系统提供的数据库是其它系统所没有的。这种“只此一家,别无分店”的作法,迫使用户非用本系统不可。另一种情况是,几个系统都装载同一数据库,但有的系统只能装载题录,而只有一家系统才既有题录又有摘要,因而以记录的信息含量丰富取得支配的地位。例如前面提到的《化学文摘》数据库,仅 STN 系统才能提供包含摘要的全记录。在这里,那些本身既是数据库的生产者,又是联机检索系统的经营者占有有利的地位。也正因为如此,一些数据库生产者近年来纷纷自己经营起联机检索系统来。

有的系统将使用量密集(如商业、医学等)的一批数据库组成数据库群,提供“一揽子检索”(one search)的功能,即允许用户对群中的数据库实行一次进入、自动转库、并自动清除被检索出的重复记录,从而使用户得到很大方便。

有的系统以用户界面的改进作为竞争的手段,致力于简化检索命令,使检索语句的表达更接近自然语言,从而以易用性争取那些对检索不具备很多经验的最终用户。

有的系统则以降低检索费用作为手段以发展用户量。例如 BRS 系统以检索的价廉取胜。不少系统开办了晚间和周末服务,以低价提供检索。

对用户进行培训,也是当今联机检索系统的竞争手段之一。系统设置培训文档,推出模拟教学系统,举办讲座,赞助用户协会,编印各种宣传材料,参加各种学术会议并在会议期间进行系统展示等等,这些争取情报

用户的措施是有效的。从用户培训入手,是目光长远的战略性措施。

有的系统延长作业时间,以适应世界各地用户,克服地理时差的影响,争取国外用户。

总之,联机情报检索系统之间的相互竞争是激烈的。这种竞争受信息市场规律支配,同时也促进检索系统性能的迅速提高,检索的单位费用的下降,系统的易用性提高。

近年来,各联机情报检索系统之间的竞争产生了新的格局。早期那种以装载几乎相同的一批数据库而互相竞争的情况减少了,而代之以强调形成各系统自己的数据库群特色,使检索系统的学科领域与用户对象趋于专业化。其次,各检索系统之间设立网关,互相沟通,改变了早期各系统之间的以邻为壑的状况。这种网关的设立,是与各系统趋于专业化的变化分不开的。网关的设立,不仅不会产生系统间数据库的重复冲突;而且有利于用户延伸所能检到的数据库的范围,从而增加了系统的竞争力。

随着各国联机情报检索系统的增多,联机系统在地理分布上发生了变化。继美国之后,欧洲众多的联机系统也建立和发展起来。随着专业性检索系统的增多,打破了原先综合性联机系统的垄断局面。各联机系统的用户正趋向分散化和均匀化。目前,联机情报检索系统经营者正面临光盘检索系统(集光盘数据库与检索软件于一身)经营者的挑战。光盘检索系统经营者已推出数以千计的光盘数据库产品(1992年报道有2021种),光盘检索软件的多窗口、多色彩的用户界面更为吸引人,检索功能更为灵活多样。尽管对光盘系统和联机系统的全面性能与费用的比较尚有争议,但一个事实是光盘系统已争得了大量的检索用户。这是联机情报检索系统面临的外部竞争。联机系统经营者在这场方兴未艾的外部竞争面前,也已开始实行了变通策略,例如DIALOG系统已发行了数十种光盘系统,并且其数量还在迅速增加。其它系统也在仿效这种作法。

四、几点启示与思考

计算机联机情报工业虽然只有20多年的历史,但科学技术的迅速发展,不仅为其提供了不断更新的技术条件,使系统性能得到迅速的改善,也为联机情报工业开辟了广阔的情报用户市场。从总体上来说,正是科学技术造就了计算机情报检索工业,并且成为它的支柱。科学技术是第一生产力,在这个领域表现得很明显。但从根本上来说,计算机情报检索工业的繁荣,必须以社会的较高情报需求为前提。而这种需求又来自社会的较高的经济与文化科学发展水平。计算机情报检索工业首先在发达国家兴起,正是它根植于发达的经济与科学文化的条件。从整体上来说,经济发展的水平决定了这个工业的水平。

计算机情报检索工业的形成与发展,是情报检索从科学文化事业转变成商业性经营的过程。事实表明,尽管在其发展初期,政府给予了扶植,使其启动,但是产业化的道路和企业经营机制却是它本身不断得到发展壮大的内在因素。只有这样,才能使它密切适应信息市场的需要,经历竞争,承受压力而迸发出活力。信息市场的动向正是社会情报需求变化的反映,以市场为引导,也就是以满足社会的情报需求作为经营的出发点。而经济效益成为经营的生命线,促进了这个工业部门的维持与发展,使其本身具有“造血”功能。

数据库的生产是计算机情报检索工业的先导和基础。从宏观来说,检索系统的建立是对数据库潜在价值的开发。一个国家要形成发展自己的计算机情报检索工业,必须首先发展数据库的生产。没有这个基础,就等于无米之炊。仅仅依靠从外国引进数据库,那么在很大程度上就会受制于人,使这种工业失去自立自强的基点。

商业与经济领域是计算机情报检索工业的主要服务对象和销售市场。在早期,计算机情报检索是以为科学技术研究服务为中心的,数据库的内容以揭示世界科学技术文献为主,数据库在类型上以书目数据库为主,在服务对象上以科学研究人员为主。但是,在70年代以后,特别是80年代以来,由于商业与经济部门信息需求的增长,而且这些部门把及时准确信息的获取作为企业运行与管理的一个必不可少的因素,信息的投入能够实现更多的产出,信息直接转化为利润,因而这些部门成为联机情报检索系统的主要用户。与此相适应,计算机情报检索工业也实现了数据库与用户对象的重点转移。这表现在:经济与商情数据库在发展速度甚至在绝对量上超过了科技文献数据库,服务对象从学者扩大到商人与一般公众,在检索方法的设计方面更多地面向最终用户,在数据库类型方面则源数据库增长迅速。这种变化看来是带有普遍性的,在美国是如此,在西欧也是如此。

计算机情报检索工业要形成规模,需要有图书情报部门、新闻出版公司、计算机分时公司及其他拥有雄厚资金的投资者的参与。这些部门构成了这方面工业的主体。除少数情况之外,一般图书馆仅仅是中间用户,作为最终用户与联机情报检索系统之间的桥梁。计算机情报检索工业也象其它工业部门一样,没有投资,就形成不了“气候”。没有充足财力的支持,就难于不断更新技术设备,扩大经营范围,顶住经营的风险,实现自身的发展。

从总体来看,我国计算机情报检索仍停留在由国家制订规划、靠政府拨款维持的事业性质阶段上,基本上未形成一种工业。虽然近年来,许多情报单位的计算机情报检索开始收费,实行有偿服务,但是从根本上来说,仍缺乏作为产业经营的经济运行机制,缺乏企业界的投资,缺乏出版界、新闻社等方面的参与,缺乏竞争,缺乏自身造血功能。这种状况,使我国计算机情报检索的发展步履艰难。我国作为一个发展中国家,情报用户的情报意识和需求水平较低,数据库编制与软件研制的基础较差,电讯设施不足且费用较高,原始文献提供的能力不足,等等,这些都决定我国短期内不可能形成如同发达国家那样水平的信息市场和计算机情报检索工业。

但是,我国在发展计算机情报检索工业方面也有优势。

首先是我国是个大国,尽管情报用户的人均检索需求水平很低微,但是全国的情报用户及其检索需求的绝对量是大的。改革开放的形势将进一步激发这种需求。特别是涉外经济活动的增多和科技查新制度的建立,社会的情报检索需求今后还会有较大的提高。

社会主义条件下的商品经济的发展,提高了人们把情报作为一种商品的意识,知识产权制度也已确立,有利于信息市场的形成。事实上,计算机情报检索工业在我国已有萌芽,国内情报检索服务已有相当数量从脱机系统转化为联机系统。社会化的信息咨询业也有发展。

我国文献资源相当丰富,这是别的国家所不能取代的。因此,我国数据库生产,经济成本低,而且内容有特色。发展我国的数据库生产业,既能建立我国自己的计算机情报检索工业的基础,又能体现我国“情报自立”的精神。

根据我国的国情,计算机情报检索工业将以国家情报机构为主,使其逐步走上企业化经营的轨道。同时也有必要鼓励新闻出版界和其他企业界在这个方面进行投资。

参 考 文 献

- 1、Conveg. John; Online Information Retrieval, An introductory manual to principles and practice, 3rd Edition, 1989, ISBN 0—85157—438—6
- 2、Mastroddi. Franco; Electronic Publishing; The New Way to Communicate. Proceeding of the symposium held in Luxembourg, 5—7 Nov. 1986 ISBN 1—85091—263—7

(本文责任编辑 江 平)