

联机检索入门

国际联机检索部 译

ONLINE SEARCHING

An Introduction

W.M.Henrg

Butlerworths, 1980

联机检索入门

国际联机检索部译

科学技术文献出版社出版

1202印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

开本: 787×1092¹/₃₂ 印张: 10.25 字数: 219千字

1986年8月北京第一版第一次印刷

印数: 1—9,000册

科技新书目: 124—45

统一书号: 17176·487 定价: 2.10元

目 录

第一章 概论	(1)
第二章 联机检索服务与数据库的起源和范围	(8)
2.1 一种工业的诞生	(8)
2.2 提供联机检索服务的机构	(10)
2.3 主题范围	(13)
2.4 数据库的规模	(19)
2.5 文献描述	(20)
2.5.1 文献记录	(20)
2.5.2 打印格式	(24)
2.6 数据库——值得注意的几点	(28)
第三章 硬件	(32)
3.1 引言	(32)
3.2 网络	(33)
3.3 电话	(35)
3.4 计算机终端及其选择	(36)
3.5 调制解调器	(41)
第四章 文献检索方法	(44)
4.1 主要操作	(44)
4.2 接通联机系统	(45)
4.3 选择准备检索的文档	(48)
4.4 具体检索——选择和组配检索词	(50)
4.5 打印结果	(53)

4.6 若干分系统的特征	(56)
4.6.1 改错	(56)
4.6.2 在终端上取得关于系统的信息	(58)
4.6.3 传送信息	(59)
4.6.4 检索词的分级列表	(59)
4.6.5 列表一次检索细节	(59)
4.6.6 删除文献组或检索语句	(60)
4.6.7 转换文档	(61)
4.6.8 保存检索策略	(61)
4.6.9 取得计时清单	(63)
4.6.10 订购原始文献	(63)
4.7 取得有关联机检索的情报	(63)
第五章 数据库结构	(67)
5.1 综述	(67)
5.2 记录结构	(69)
5.3 文档结构	(70)
5.4 加一个记录到数据库中	(74)
5.5 字典	(74)
5.6 词组的处理	(82)
5.6.1 标记词组法	(82)
5.6.2 匹配词组法	(84)
5.7 文献记录倒排	(84)
第六章 实用检索方法	(88)
6.1 引言	(88)
6.2 单词	(88)
6.3 词的截断	(89)

6.4 显示词典	(92)
6.5 字段检索	(95)
6.6 词组检索	(99)
6.7 组配词与检索语句	(100)
6.8 位置运算符	(103)
6.9 字符串检索	(104)
6.10 限制检索	(106)
6.11 检索保留方法	(107)
6.12 多文档检索	(108)
6.13 检索重新显示	(109)
6.14 通用指令语言	(109)
6.15 小结	(110)
第七章 检索准备和检索策略	(113)
7.1 典型的联机检索	(113)
7.2 进行准备	(115)
7.3 接受检索咨询	(115)
7.4 确定咨询课题是否适于联机检索	(117)
7.5 选择哪些文档进行检索	(118)
7.6 使用哪些系统	(119)
7.7 分析检索课题	(119)
7.8 检索词的选择	(122)
7.8.1 检索受控标引词	(124)
7.8.2 检索非控制词	(128)
7.8.3 检索作者	(130)
7.8.4 检索高命中数的词	(131)
7.8.5 引文检索	(132)

7.8.6 选择检索词——一个例子	(132)
7.9 准备计划	(134)
7.10 在终端上	(135)
7.11 获得结果	(136)
7.12 化学品的检索	(138)
第八章 联机检索员的作用	(142)
8.1 引言	(142)
8.2 各种检索者的比较	(144)
8.3 联机检索员的职能	(146)
8.3.1 智力职能	(146)
8.3.2 管理职能	(147)
8.4 联机检索员应具备的素质	(147)
8.5 小结	(149)
第九章 联机检索的管理工作	(152)
9.1 管理工作的作用	(152)
9.2 向上级管理部门进行说服工作	(152)
9.3 为联机服务工作进行宣传	(154)
9.4 人员培训	(155)
9.5 服务工作的建立	(156)
9.6 服务工作的管理	(158)
9.7 费用管理	(159)
9.8 收取服务费	(160)
9.9 对工作人员的影响	(162)
9.10 开展联机检索的任务	(163)
第十章 教育和培训	(166)
10.1 引言	(166)

10.2	谁来组织教育和培训工作?	(167)
10.2.1	教学机构	(167)
10.2.2	数据库的编制单位	(168)
10.2.3	单独的机构	(168)
10.2.4	国际机构	(169)
10.2.5	国家机构	(169)
10.2.6	专业学(协)会	(170)
10.2.7	操作系统	(170)
10.2.8	用户团体	(170)
10.3	课程设置	(171)
10.3.1	综述	(172)
10.3.2	专业介绍	(172)
10.3.3	示范检索	(174)
10.3.4	实际检索	(174)
10.3.5	检索结果的评价	(175)
10.4	教具	(175)
10.4.1	声频记录和数字记录	(177)
10.4.2	计算机辅助教学	(179)
10.4.3	打印材料	(180)
10.4.4	其它的手段	(181)
10.5	小结	(183)
	附录说明	(187)
附录 1	检索准备和检索策略的核对表	(188)
附录 2	Blaise	(195)
附录 3	ESA-IRS	(202)
附录 4	Infoline	(216)

附录 5	Lockheed	(222)
附录 6	SDC	(267)
附录 7	Euronet	(295)

第一章 概 论

目前，用简单的打电话方式，就能在数以千万计文献文摘中检索出几百条有关信息。人们可以在相当于一百万册或更多的书籍范围内进行检索，从而找到与一个特定课题有关的文献目录。联机检索所需要的全部设备，除了电话机以外，另有一台电传打字机和一个将其连接到电话机上的装置。电传打字机与打字机十分相似，只不过它有与计算机通讯用的附加键。接线盒的大小和这本书差不多，它把来自电传打字机的电讯信号转变成声频信号，并沿电话线发送出去，反回信号的传输过程同样如此。由于这些系统简单，并有大量情报可供利用，因此在检索情报方法上已经产生了明显的效果。

近年来，技术情报一直在急剧增加，以致大多数人跟不上它的发展步伐，甚至都跟不上其本身专业领域里知识的发展步伐。由于期刊数量的显著增长，促使二次文摘服务工作加速发展，这种服务能使用户在查阅特定课题时，得到有关文献的原始刊登期刊，正是这些文摘期刊所包括的大量资料才间接地缓和了文献检索的矛盾，至少在一段时间内暂时缓和。

计算机特别适合于解决在尽可能短的时间内，准确地把收藏的文摘中参考文献打印出目录表。当信息为电子形式时，修正与编辑的过程是直截了当的，存入计算机内的情报

能够用与计算机连接的键盘进行核对，需要时也可加以修改。同样，当信息以电子形式传递时，需要不同形式的文摘资料，例如以字母顺序排列的作者表，以关键词字顺排列的倒排档，以分类顺序排列的分类表，以及其它种类的列表，信息的排序和重新编制格式都要简便得多。此外，可以把存有信息的磁带送往打印机，直接生产印刷版本的二次文献。电子计算机程序引入的信号可以制出各种不同的字号、字型 and 排版格式，并且都能做到原文照排。因此作为经济有效的出版刊物的副产品，以计算机可读形式给出期刊目录的完整清单，作为计算机检索的成品供用户或单独机构使用。

首先使用该技术的机构之一是美国国立医学图书馆，它在六十年代中期生产出印刷版本的 Index Medicus 以及相关的 MEDLARS (Medical Literature Analysis and Retrieval System) 文档。从那时起，许多机构一直利用计算机出版文摘期刊，因而形成了机读的书目文档，这些文档通常称为目录式数据库。目前可以使用的目录式数据库范围包括大部分科学技术学科的参考文献，并且还逐渐增加了教育学、社会学、经济学、商业和人文学等领域的参考文献。此外，现在有许多数值信息的机器可读文档，这些文档通常称为数值型数据库 (Databanks)，同样还有一些包括文献全文的情报系统，可供检索。

在六十年代末期，这些数据库才开始以批量检索服务的方式提供用户使用。所谓批量检索服务就是将查询的问题提交中心机构，然后加以汇总。做到以有效利用费用的方式来检索计算机文档，只是检索者要等一些时候才能得到检索结果。然而这种提供信息的方法对公众的影响不太大。近来计

算机情报服务的急剧发展是取决于计算机信息交换的通讯网络的广度和可靠性。这些发展首先是由银行、建筑学会和大的公司这样一些大机构兴起的，因为对这些机构中的许多单位来说，得到准确的现行数据是很重要的。这样发展起来的技术，导致发展通讯网络，使任何人都可通过一般的公用电话网络使用计算机。

七十年代初期，北美开始使用Tymnet和Telenet网络，不久欧洲以及世界其它国家都开始使用这两个网络。Euro-net网是提供传输数字信息的欧洲网络。

另一重要的发展是建立了一些机构，它们提供了必需的计算设备以存储这些数据库，并使检索者以联机方式直接与数据库联系。像这样的机构有美国的洛克希德公司 Lockheed、系统发展公司 (System Development Corporation 简称SDC)、文献目录检索服务处 (Bibliographic Retrieval Service 简称BRS)、英国图书馆自动化情报服务处 (British Library Automated Information Service 简称Blaise)、Infoline系统，以及欧洲的德国医学文献情报研究所 (Deutsches Institut für Medizinische Dokumentation und Information 简称DI MDI) 和设在意大利的欧洲空间组织 (European Space Agency 简称ESA)。因而，通过使用公共电话网络和大量以计算机可读形式出现的参考文献库的结合，形成了当今情报检索的简捷方法。

这些发展促使检索量有了惊人的增长，威廉姆¹ (Williams) 估计：1974年检索量有70万条，1975年为100万条，1976年为120万条，1977年为200万条。英国的检索量一直在迅速增长，1978年已超过10万条。

联机服务的主要用户来自工商界，其次是学术机构。公共图书馆使用该业务的量不大，但正在增长。

人们通过寻找要检索的词，即寻找与以期刊文章里叙词形式存储于数据库内匹配的词进行检索。这些检索词可以是单词，例如：Information, Retrieval等，也可以是短语，例如Information retrieval或Library management systems。当你确定不了准确的词时，也可检索该词的截断。例如 Comput? 将给出以这六个字母为词首的一切词和词组，如Computer, Computing, Computable, Compute, Computation, Computer-aided design, Computer system等等。同样ON*LINE将和online, on-line, on line 是同一类短语。在某些系统中，还有一些方法允许检索一个词的片断，这对检索化学名称特别有用。可将这些检索词限制为有效的规范词表中的词，标引者和检索者必须使用规范词；另一种方法是，可以使用自由词。检索者在使用时，可以用任何描述检索题目的词。也可以从许多字段，例如题目、文摘、标引字段里的词中寻找相匹配的词，或者检索一特定的作者姓名或期刊刊名，或者作者所属机构名称。在某些主题范围，有可供检索的标准分类代码。上述种种可能性提供了非常灵活的检索系统，各单个的检索词能够与特定几个词中任何一个必须出现的词进行组合，也可以与它们必须一起出现的词进行组合。

典型的检索过程总是从检索者选择检索词开始的，选择检索词有两种方法，一是查阅书本式的经审定的检索词字典，二是查看终端上打印出来的、储存在计算机内的字典。然后，把选出来的检索词送进计算机，每输入一个检索词，

计算机就打印出该词找到的文献篇数，用户就可把各个检索词进行组配，或试用新的检索词以修改不合适的检索策略，直到从数据库中选出的篇数合适为止。在打印资料量小的情况下，可以在联机终端上打印选定的文献，否则就在计算机所在地脱机打印，打印件通过邮局迅速航寄给用户。通常能够以各种不同格式打印所选的文献：最简短的格式，如仅有一个文摘号；或文摘号加标题；或完整的全记录，包括：作者、标题、文献出处、分类代码、标引词，完整的文献全记录另加一篇大约600字的文摘。目前使用的系统一般不能检索文献全文，然而有些系统允许用户从终端上订购原文的照相复制件或缩微胶片复制件。

刚开始用联机检索的人是愿意了解计算机检索系统与常规手工检索文献的方法相比较有什么异同的。计算机系统最显著的特点是每一次检索都要计费，使得情报费用成为一笔可观的支出。倘若一位研究助理在图书馆花费几天时间翻阅参考资料，那么就不易看到一个机构所付出的费用。人们很习惯于接受不易觉察的支出，而不愿接受明显的支出。因此，联机检索的费用会是决定使用系统的一个重要因素。对新的用户来说，会迅速明白联机系统的能力的。它可以简便地把广泛的同义词列表，并与其它相似的文献组进行组配来完成一次检索，这种复杂的检索用手工检索就会令人非常恼火。机检时，系统立即显示出：用户选定的检索策略所查到的文献篇数，可为用户反复改进检索策略提供方便，直到获得满意结果为止。最后用户能得到一份打印的文献目录单，这种方便是不会被任何一位辛辛苦苦从文摘期刊上抄下几十篇文献目录的先生低估的。典型的联机检索一般能使检索者直接

检索文献记录的各个部分，以及标题或文摘中的单个词，从而减少使用印刷版索引时所必需的许多检索步骤。联机检索的另一个优点是速度快。联机检索比手工检索同一文献库的速度要快许多倍。约翰逊² (Johnson) 在一次研究中发现，联机检索平均只需手工检索所化时间的六分之一。

当然，联机检索也有一些问题。一是需要专门设备，另一个是需要合适的人来进行检索。在第八章里，将对此作进一步论述。在进行一次检索之前，偶尔还会因为设备出现故障耽误一些时间。目前，大多数数据库只包括大约从1970年以来出版的文献，而且对大部分主题来说，还不能检索1970年以前的文献。手工检索虽然有时也会同样有效（例如查阅已知作者的论文，或索取特定主题标题的文献样本），但是，一般来说，使用系统的主要阻碍因素在于情报价值是否会超过取得情报的费用。

下面各章节将详细论述检索过程，并探讨联机检索的实质。从用户的反应中清楚地看到，这些新系统鼓励着那些不用联机检索就消息闭塞的许多人去联机检索文献。使用联机系统主要不是说这些系统是目前我们执行任务的一种有效利用成本的方法，而是说新的花费将会使普通的用户在接触其学科领域的情报方面达到一个完全崭新的水平。在过去，获取丰富的能源曾给我们的生产力带来了惊人的变化，那么现在计算机的本领也做到了这一点。也许，容易得到累积的情报会导致另一次人类成就的急剧增长。

参 考 文 献

1. Williams, M. E. 'Database and online statistics', *Bulletin of American Society of Information Scientists*, 4, No. 2, 21-23 (1977)
2. Johnston, S. M. 'Choosing between manual and online searching—practical experience in the Ministry of Agriculture, Fisheries and Foods', *Aslib Proceedings*, 30, Nos 10, 11, 383-393 (1978)
3. Atherton, p. and Christian, R. W. *Librarians and Online Services*, Knowledge Industry Publications, White Plains, N. Y. (1977). ISBN 0-914236-13-X
4. Cuadra, C. 'Commercially funded online retrieval services—past, present, and future', *Aslib Proceedings* 30, No. 1, 2-15 (1978)
5. Hall, J. L. *Online Information Retrieval Sourcebook*, Aslib, London (1977). ISBN 0-85142-106-7
6. Hawkins, D. T. 'Online information retrieval bibliography'. Second update, *Online Review*, 3, No. 1, 37-73 (1979)
7. Houghton, B. and Convey, J. *Online Information Retrieval Systems*, Bingley, London (1977). ISBN 0-208-01660-0

第二章 联机检索服务与数据库 的起源和范围

2.1 一种工业的诞生

由于三个主要因素，致使我们今天能这样联机检索服务。

第一个因素，有一些组织、公司、团体等，专门汇编和审核收藏的文献目录。虽然大多数这种馆藏目录原来打算编制印刷版索引的，但是近几年来，已经越来越多地把它们制成机器可读版，从而使得生产者已经能够利用计算机来编制索引。一旦采用这种机读版，当然，它就高度适宜于用计算机进行更加广泛的处理，而且，如果联机检索服务要不是及时提供这些馆藏资料的话，毫无疑问，联机检索服务工作一开始就会受影响，致使发展迟缓。即使现在，把过去某个时期的资料追溯转换成机读版花钱还是很多的，并且难以证明这样做是否合算。

第二个因素，是通讯和计算技术的最新进展。有了这种技术进展，才能建立起可靠的远程通讯网络，并且把那些带有强大数据处理能力的计算机和联机储存能力大的计算机连入网内。这样组合起来的计算机网络是如此有效，以至于现在无论是在房间的另一侧还是在世界的另一端对机器进行

提问，简直就没有任何明显的差别。

第三个因素，情报资源已经和新技术紧密地结合在一起了，并把这种相结合的产物从一种设想，精心培育成为现实的专业知识和企业单位。它们主要任务是从对应的原始情报着手，为建立和管理联机检索用的高度结构数据库而编制了各种程序（软件）。这些软件还必须保证用户避免复杂的计算机操作，允许用户以相当简单的方式与系统“对话”。众所周知，“用户友谊”这个概念是非常重要的，因为它会有力地影响着用户对系统的态度。

把情报资源和新技术结合在一起的单位有各种各样的名称：主机、主机操作、主机系统以及本书称为操作系统等等。显然当前的主机运行经济学认为不是所有名称都会继续存在下去，但是名称的数目仍在慢慢地增加。联机系统正在利用印刷版索引工业的副产品，这个事实提出一个有趣的问题。在印刷版索引工业保持健康发展时，这个问题并不大，但是，假如订户不断转向操作系统为他们的情报联机做好准备，那么，两种服务的任何一种要避免脱节，都不得不对价格作一些灵活调整，以求经济平衡。

向来以因循守旧而著称的图书馆和情报界竟然靠这种先进技术成果获得解放，实在令人啼笑皆非。计算机应用问题，特别在联机检索方面，并不是没有受过图书馆和情报界的挑剔，过去十年曾经目睹情报领域的复兴，其中又以联机检索作为推动变化的主力，这一点现在几乎没有什么人还会有异议。由于通讯和计算技术继续迅速发展，毫无疑问，即使现在还认为了不起的能力，在八十年代过去之前将会变得微不足道。