

温有奎 编译

国际联机与光盘检索 策 略



陕西科学技术出版社

内 容 简 介

本书是选择国外长期从事联机情报检索教学与服务的著名专家们的优秀专著编译而成的。对联机情报检索的原理、过程，特别是查找情报的方法与策略，数据库建设及相应软件接口，光盘检索与联机检索的比较有精辟论述。

本书可供情报学系（专业）、高校理、工、农、医、文各专业开设的“文献检索与利用”课的师生，各级情报单位的情报工作者使用。实践证明，对广大科技工作者迅速获取研究课题中的对口技术资料，及对取得科研成果鉴定的技术资料证明有极大帮助，深受科技工作者欢迎。

前 言

国际联机情报检索和光盘检索已成为获取世界科技和经济情报必不可少的重要工具和手段。自 1983 年 3 月，我国开展国际联机情报检索服务以来，全国设置的国际终端已达 100 多个，遍及 26 个省、市、自治区的 47 个城市，开通的国际联机情报检索系统已达 12 个，检索课题已超过 10 万个。光盘检索服务出现较晚。两年来，我国已有 100 多个单位租用国外光盘进行检索服务，光盘检索也受到我国用户的欢迎，用户的人数飞速增长。目前这两大计算机检索手段已成为我国科学技术、经济、贸易等各个领域获取世界情报的最快、最新的工具。

联机情报检索为我国的四化建设发挥了重要作用。国际联机情报检索吸引了我国的广大用户，但也吓走了不少用户，以致至今人们对国际联机检索还感到高深莫测。国际联机检索使用得当，花费不多，能及时取得世界最新情报，了解或掌握世界科技、经济等领域的最新动态；用的不当，则花费惊人，还得不到满意的结果。相比之下，光盘检索费用低，对检索策略的要求不高，一个课题可反复查找。但两年来的检索实践表明，有时同一课题，光盘检索的效果没有国际联机检索的效果好。往往一个课题多次上机，却检不出用户真正需要的情报时，用户就会对光盘检索失去热情。归根结底，国际联机情报检索和光盘检索都需要具有较高的检索

策略。

在从事国际联机情报检索和光盘检索服务与教学工作中，我深感，提高我国科技情报工作者机检的业务水平和向广大用户宣传介绍联机检索和光盘检索的基本知识已迫在眉睫，在全国著名机检专家、武汉大学图书情报学院情报研究所所长陈光祚教授的热情指导下，我根据个人的工作实践，参考了国外长期从事联机情报检索服务和教学的著名专家们的著作，编写了这本书。本书旨在对联机检索的原理、过程，特别是对查找情报的方法与策略给出精辟的论述，并且重在提出取得最佳检索结果的各种模型。作者期望本书对于我国广大科技情报工作者、图书情报学专业的师生、学习《文献检索与利用》课的学生、利用计算机获取情报的用户能有借鉴和指导作用。本书若能对我国国际联机情报检索和光盘检索的水平有所促进，作者将感到不胜欣慰。

由于水平所限，错误在所难免，敬请广大读者及机检专家随时指正。

编译者

于西安电子科技大学

1990年5月

目 录

第 1 章 人一机交互的能力与局限	(1)
第 1.1 节 步骤与策略	(3)
第 1.2 节 计算技术与认知能力	(5)
第 1.3 节 时空观念的改变	(12)
第 2 章 联机检索中的数据与情报	(17)
第 2.1 节 数据库组织	(19)
2.1.1 书目模型	(19)
2.1.2 主题检索	(24)
第 2.2 节 自然逻辑和形式逻辑	(25)
第 2.3 节 书目目录和数据库	(31)
第 2.4 节 用户研究	(34)
第 2.5 节 联机检索	(35)
小结	(40)
第 3 章 国际联机检索系统	(41)
第 3.1 节 DIALOG 系统	(44)
第 3.2 节 BRS 系统	(46)
第 3.3 节 ORBIT 系统	(48)
第 3.4 节 ESA-IRS 系统	(49)
第 3.5 节 STN 系统	(51)

第4章 检索基础	(60)
第4.1节 协议和指令	(61)
4.1.1 建立集合	(61)
4.1.2 关键词, 词组和字段限定	(64)
4.1.3 受控词汇与主题代码	(70)
第4.2节 组配集合	(74)
第4.3节 显示和打印	(81)
4.3.1 基础	(81)
4.3.2 策略性显示	(82)
4.3.3 联机打印和脱机打印	(86)
4.3.4 检索举例	(87)
第5章 检索策略	(98)
第5.1节 概率因素	(100)
第5.2节 集合代数和组配阵列	(105)
小结	(124)
第6章 检索原理	(126)
第6.1节 表达检索	(126)
第6.2节 语义检索与语法检索	(130)
第6.3节 定性检索与定量检索	(162)
第6.4节 完成检索	(165)
小结	(168)
第7章 选择数据库	(169)
第7.1节 认识上的限制	(170)
第7.2节 三次数据库的困境	(172)
第7.3节 过库检索和知识增长	(177)

第 8 章 心理因素	(179)
第 8.1 节 用户的智力模型	(179)
第 8.2 节 训练与经验	(183)
第 8.3 节 人的能力和人的局限	(186)
第 8.4 节 关于 NOT(非)问题	(194)
第 8.5 节 问题求解	(195)
小结	(196)
第 9 章 检索辅助软件	(197)
第 9.1 节 用户友好概念	(197)
第 9.2 节 软件接口、入口和前端	(199)
第 9.3 节 检索辅助设备比较	(202)
第 9.4 节 远程系统前端	(212)
第 9.5 节 未来的辅助检索	(217)
第 9.6 节 第五代计算机: 有思想的机器	(219)
第 9.7 节 计算机辅助教学和图解显示	(222)
小结	(225)
第 10 章 数据库结构和工作站	(226)
第 10.1 节 熟悉计算机	(227)
第 10.2 节 微机硬件	(229)
第 10.3 节 操作系统	(232)
第 10.4 节 访问软件	(235)
第 10.5 节 通讯	(244)
第 10.6 节 管理文献的广义策略	(250)
第 10.7 节 硬件建议	(264)
第 11 章 光盘情报检索	(255)
第 11.1 节 光盘检索进展	(255)

第 11.2 节	联机检索系统功能	(258)
11.2.1	DIALOG 系统常用指令及用法	(259)
11.2.2	索引	(263)
11.2.3	国际联机检索系统的功能	(265)
第 11.3 节	光盘检索系统的功能	(268)
11.3.1	光盘检索系统常用指令及用法	(268)
11.3.2	系统查找特点	(268)
11.3.3	基本查找	(271)
11.3.4	构造复杂提问	(272)
11.3.5	优化查找	(275)
	小结	(277)
附表		(278)

第 1 章 人一机交互的能力与局限

由于观察问题的角度（或出发点）不同，不同的用户对同一个计算机系统的软件接口会作出不同的甚至是截然相反的评价。有的认为，它容易使用，有的则认为它难于使用。例如，可能有些专家会认为本来一个“容易使用”或“用户友好”的系统是蹩脚的；因为，用这种系统完成一项任务时，通常要经历由菜单驱动的许多步骤。而初学者则会认为这个系统对他们很有帮助，因而欢迎“用户友好”接口。相反，专家会由于一个系统的使用效率高而认为其容易使用，但对初学者来说，则很可能认为难于学习，望而生畏。

人们对计算机有许多不同的看法。这些看法大多涉及计算机与人交互的能力，如由计算机引起的技术革命对社会产生的影响。但是，最好的还是说明用户与计算机之间的关系。例如：

1. 计算机越容易使用，它的灵活性就越小。
2. 计算机的灵活性越大，使用起来就越困难。

这两句话阐明了人们通常用计算机可能做什么，而不是人们通常做什么。下面，把用户和计算机的上述关系稍加修改，会引出计算机的使用率问题。

1. 计算机越是容易使用，其被使用的可能性就越大。
2. 计算机越是难以使用，则其被使用的可能性就越小。

因此，易用性是设计和选择计算机系统时的一个非常重

要的问题。用户不仅必须对灵活性和用户友好作某种权衡，还必须对选择或使用计算机系统作出权衡。换句话说，系统的灵活性与简单性之间存在着互逆关系。

本书的方法之一，是把在一定程度上作相同工作，但使用的难易程度有差别的系统作了比较。也就是从情报检索的角度，考察具体系统的类型。但是，其灵活性和简单性的关系同样适用于程序语言和文字处理，甚至游戏程序在内的所有计算机领域。

数据库管理和联机查找系统对易用性和能力范围都作了很明显的权衡。一个特别好的例子是科学情报协会（ISI）的 Sci-Mate 系统，该系统由三个不同的应用软件包组成：查找包，管理包，编辑包。

这三个软件包组合在一起时，可以协同工作，否则，也可以单独使用。查找包是前端机和入口的组合。前端机为查找远程数据库系统，提供用户友好菜单接口，使用同一前端机软件的入口，能提供访问不同的国际系统，如 DIALOG 等。访问是通过向前端机的菜单接口提问来实现的，接口把用户的询问转换成查找系统的命令语言形式。

接着，使用管理包存储查找所得的检索结果，这样用户能够定期地查找和更新个人数据库。系统的第三个部分是编辑包，它把来自个人数据库中的检索结果按特定的书目形式格式化，必要时还可改变书目形式。

该软件使用起来很容易，而且相当灵活，但比不上更普通的应用包那样灵活。例如，通过安排自动进入检索系统的通讯，用户只须按下一个键就能招唤一个“宏指令”，将进入远程系统要求的二、三十个字符一次发送出去，但是这种宏

指令“(macro)”只在五个检索中心中使用。

另一种软件包，如 Smartcom（这是为建立通讯专门设计的），允许用户可对任何一个系统写出一个宏指令。选择通讯软件时，对这两个系统必须进行易用性和灵活性权衡。Sci-Mate 使用起来比 Smartcom 稍容易些，但通讯参数选择的灵活性则不如 Smartcom。

把 Smartcom 与其它通讯包，如 Kermit，做比较，更能说明这一问题。Smartcom 只适用于某些特殊的终端，其仿真数目有限。Kermit 则允许用户规定独特的参数，具有建立各种终端的仿真能力。然而，Kermit 比 Smartcom 语义更含糊，使用更简练。于是，Sci-Mate，Smartcom 和 Kermit 比较，易用性依次降低，而能力和灵活性则依次增加。

上述这些系统，在现有信息处理系统中都是比较常见和比较流行的。用它们能对远程数据库中联机检索得到的记录进行访问、存储、处理和编辑。

数据库的概念特别重要：除了文字处理之外，无论在信息的获取还是在信息的管理方面，它都是专业人员应用计算机的最重要概念。不仅如此，对于使用计算机的用户来说，由于使用数据库查找要涉及到策略问题，因此数据库对他们也是最重要的概念。用计算机进行交互有两种类型，一种是策略性的（如对数据库的），另一种是步骤性的。

第 1.1 节 步骤与策略

步骤是涉及特定的原因和效果的活动。文字处理是步骤

活动的一个极好例子。要改变文本中的一个词（比如说，将“这个”改换成“那个”），用户只要给出适当的命令就可实现。这一步骤是简单直观的，用户只要知道进行这种改变的适当命令就行了。这时，策略用的非常少。用户要完成的任务非常明确，即把“这个”改成“那个”。

由于查找数据库需要一定的策略，这就使得数据库的使用变成一种极为复杂的活动。用户访问数据库是因为他们想从中得到情报。如果得到的情报未满足要求，他们将不得不再追加提问信息并重复这个过程。这里，用户是在与未知打交道：他们有一种情报需求，并力图从一个数据库系统中满足这种需求，而这个数据库并未能提供所需要的全部情报或者只提供部分情报。用户不仅要对未使用的策略作出判断，而且要对检索的结果作出解释。

查找是否成功，首先取决于该系统中是否存有所需要的情报。但是，更重要的是用户向系统正确提问的能力。情报可能在数据库中，但如果用户不能有效地检索它，则同样不能得到它。如果所需要的情报不在数据库中，正确的检索策略能使用户明确了解它确实不存在于该数据库中。

在数据库中检索情报通常包括与系统交互对话（一种讨论，就好像一个人在寻找某一件东西时与另一个人进行对话一样）。换句话说，交互的对方都有反馈发生，且后一个提问都需要以前一个提问的答案为依据。

对话的主要差别是：在数据库中的交互要受交互的对方（计算机）的水平限制。这就是说，用户能理解计算机怎样处理数据的步骤，但计算机不能以用户的水平理解用户——

如何处理情报或如何思考。于是计算机仅仅能执行程序而

不能在策略上提供任何帮助，也不能思考，而这一点在智能交互中，是一个至关重要的因素。

因此，用户必须把他的需要转换成计算机能够识别的形式。这个转换过程包含有专门的指令。尽管这些指令为数不多，但使用起来却需要大量的智力劳动（取决于需求的复杂程度，需求越复杂，智力劳动的程度就越大）。在第 4、5、6 章中将会看到，指令的形式是足够的简单；而指令的应用则很困难。

计算领域的发展趋势以及它的整个历史，都是基于使程序性任务更容易执行这个观念。但是，当任务变成智能化的时候（例如情报检索），指令及其是否容易使用已不是所要考虑的问题，而如何使用这些指令则成为成功的关键。用户如何与系统交互，是建立在他们解决问题的技巧、认识能力和采用适当的策略知识的基础之上的。在这些“智能”领域中，计算机有很大的弱点，因而很少为用户提供什么帮助。

第 1.2 节 计算技术与认知能力

记忆能力是计算机具有强大功能的关键因素，即巨量的、精确的和无差错的记忆能力与电子速度相结合形成了解决问题的卓越能力。由于计算机的这种记忆能力和速度，它能在数秒内从上百万个词构成的词表中找出一个特定词。这种能力也是计算机具有类似人的品质或智能的原因。这些特性可能用于各种场合，例如，从一个文件中找出一个特定的词的基本操作到模拟人的思维。无论用在什么场合，成功的关键取决于计算机存有专门的指令，这些指令通过输入一条

命令或作为对用户的一个回答的响应而得以执行。

计算机指令由数据的二进制数字（即比特或位）构成。一个比特有两种状态，因而能够表示两种情况。于是一个比特开关可以是正或负，即开或关。它代表 0 或 1！比特的组合构成了计算机执行的指令。

人的神经系统也按二进制系统工作：神经细胞，即神经系统的功能单元可以处在开或关的状态。但是，最重要的差别在于表示神经和比特的网络结构的不同，特别是决定神经细胞和比特是开或关的条件不同。

决定计算机的二进制位的状态是十分简单的——由设计者和编制计算机程序的工程师和程序员来决定。不存在不确定性；某一个比特要么是开，要么是关，它的状态是预先决定好了的。因此，一台计算机所做的一切响应是预先决定好了的。一个系统在对一个用户的提问作出的响应时，可能提供多种选择，因而看起来好象它在进行思考，但实际上它能作出的决定是预先定好了的。

人的思维和认识则完全不是预先决定好了的。主观因素将影响用户如何响应，环境也是一个因素，而且人能作出的响应数量可能是无限多的。这是人与计算机两者之间的根本区别。人能思考；计算机则通过完成一系列指令来模拟人的思维。为了使计算机能与人的推理过程相比较的水平上进行思维，必须对计算机的基本单元（即比特）如何存储、连接和处理进行革命性的变革。

在当前的计算机构造中，数个单独比特在一起形成一个字节。如果单独比特是数据的基本单位，那么，由比特组成的字节则是表示一个字母或一条指令那样的信息基本单位。

正如上面所解释的，每一个比特是一个独立的实体，它的开或关状态是由计算机的设计者预先决定的。图 1.1 描述了猫 (cat) 这个词如何由一个字节来表示，并给出各字节的二进制位。

0 1 = c
1 0 = a
1 1 = t
0 0 = s

图 1.1 比特、字节、词的构造

该例子表示，一个由二比特组成的字节能够表示四个不同状态。这里使用四个状态表示字母 c, a, t, s, 拼成 cats。

实际上，字符、数字和特殊编码（如换行符）用称作字的单位表示。一个词可能由一个 4 比特的 2 字节组成，因此，8 位字长的编码将提供 256 种不同的组合，足以用来表示大小写字母、数字和各种其他表示符号。

因为每一比特是一个独立的实体，所以任一给定比特的状态是与其它比特相互独立的。它们的关系由在一个字节内的比特集体状态确定。根据开、关的组合，将显示一个给定字符或符号，或者将执行一条指令。重要的是认识到比特的状态是预先被确定的，并且各个比特的状态是相互独立的。

对于人的大脑来说，情形则恰好相反。象计算机的比特那样，神经细胞是二值性的，因为神经细胞的响应形式起因于细胞的开启 (firing) 或关闭 (换句话说，通或断)。但是，一个神经细胞是开还是关则取决于其他神经细胞的状态。神经细胞是由轴突 (axon) 和树突 (dendrite) 构成

的，轴突是通过突轴（synapse）与另外细胞通讯并发送其响应（换句话说，传播神经脉冲）的部分。树突是神经脉冲接受器。当一输入组合到这些接收器并达到一定门限时，神经细胞开启，即沿着它们各自的轴突传播一个脉冲。图 1.2 是描绘经典理论中神经细胞轴突和树突的形状。

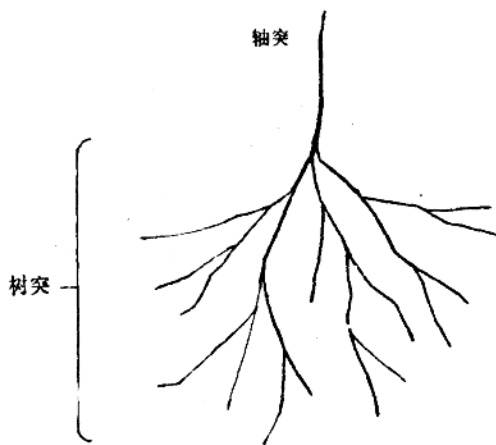


图 1.2 神经

图 1.2 中的模型是极其简单的，人的大脑有 1000 多亿个神经细胞，并且据估计，一个神经细胞可能与其它神经细胞有高达一百万个联系。由于这种极度的复杂性，人脑被称作“内部宇宙”（inneruniverse）。不仅如此，一个神经细胞处于通或断状态要由它接收的信号来确定，而且还要由存储在它的遗传物质中的信息以及突轴的传输条件（neurotransmitter conditions）决定。这些情况导致了另一

个层次上的复杂性，其复杂程度可能比神经细胞网络本身还要大。

神经脉冲是准数字的，而实际计算则是发生在神经末梢的一个模拟过程。并且，神经细胞及其树突末梢是通过一个巨大的分布式反馈网络组成的。

由此可见，就人脑的复杂程度，仅从其组成单元数目方面来说，就比计算机组成单元的数目大得多——而这种估计还没有考虑人类信息的相互连接和子细胞的条件！正是因为人脑具有潜在地产生无限响应这一复杂特性，才使人类可能得以实现推理。

推理和认知力基于类推的信息处理。二物类似。要涉及推断——即对同种的或相似的物体识别。在物理世界，这些判断包括对象音调、曲线、颜色、手势和表情等进行对比。在理论上讲，它涉及思想、语言形式和句法及语义。在所有这些领域中，变化是无穷的，集合也许在某个定量范围之内，但辨识仅是定性的。

当然，人具有通过信息处理来达到实现认知的能力，而计算机仅只限于计算。但设计人员正试图使下一代计算机——“第五代计算机”，成为能够从两个概念中产生第三个概念的推理机。换句话说，第五代计算机的能力将要广于并大于原编的程序。怎样实现这个设想尚须努力，成功是无疑的，只是时间的迟早问题。

然而，即使推理机实现了，仍然存在例如仅仅与人的常识相关的大量信息的获取问题，更不用说构成我们社会的每一个小小局部所拥有的巨量信息了。成功将逐渐到来，具有人工智能的计算机能够思维。但至少根据我们现在已知的计