

化学化工文献 国际联机检索技术

周泰应 著



科学出版社

化学化工文献 国际联机检索技术

周泰应 著

L611/L3

科学出版社

1987

内 容 简 介

本书对国际联机检索的原理、步骤及工作过程作了通俗的介绍,并结合世界上最知名的计算机检索系统——美国 DIALOG 系统,完整地介绍了联机检索系统的功能、指令及众多的数据库文档。本书通过较多实例,对 DIALOG 系统的化学化工类文档(包括 CA Search, Chemsups, CIN)的构造及联机检索技术作了详尽的说明。

读者对象:化学化工领域的科技、情报、图书等技术人员,大专院校师生。

化 学 化 工 文 献 国 际 联 机 检 索 技 术

周 泰 应 著

责任编辑 杨淑兰

科学出版社出版

北京朝阳门内大街 137 号

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

1987 年 11 月第 一 版 开本:787×1092 1/16
1987 年 11 月第一次印刷 印张:12 1/4
印数:0001—3,500 字数:280,000

统一书号:13031·3923

本社书号:5471·13—4

定价: 2.95 元

前 言

国际联机检索技术,是六十年代研制成功、七十年代已广泛应用的国际先进技术。它是情报检索、计算机、远程通讯科学技术的综合。在我国四个现代化建设中,它能配合各行各业的需要,快速提供文献情报及有关数据的线索,是一项应用面广、很有成效的技术。

本书结合 DIALOG 系统,介绍了化学化工文献国际联机检索技术。

作者根据六年来在学习和工作实践中的心得,以有限的水平,尽可能用不多的篇幅向读者通俗地介绍这项技术。作者深信,随着我国四个现代化宏图的逐步实现,这项技术必将为越来越多的地区、部门、单位和个人所采用。

本书共分六章。第一章是绪言,介绍了国际联机检索技术的一般情况、现实意义和常用术语。第二章首先以一个简例向读者展示了这项技术的全流程,随后阐述了它的基本原理、步骤和工作过程,是这项技术的主要理论基础。第三章结合国际上知名的 DIALOG 系统对这项技术的主要内容,一一作了介绍,如数据库文档、算符及语法符号、指令系列、打印格式和能力制约等。第四章在传统手检技术及了解了机用文档记录构造的基础上,以较多的实例,讨论了化学化工文献的国际联机检索技术。第五章介绍了 DIALOG 系统第二版本软件新增的及改进的功能部分。第六章阐述了联机检索产品的质量指标。

本书初稿经中国科学技术情报研究所检索室副主任兼国际联机检索服务部主任工程师何群同志审阅。作者根据她的意见,对本书初稿作了增补和修改。化工部情报研究所徐织云同志在资料方面曾给作者许多支持和方便,在此表示感谢。

最后,仅以此书纪念培育过我的母校——楚怡小学、周南女子中学和浙江大学,以及我工作多年的北京化工研究院。

作者

1986年1月15日

目 录

前言	iv
第一章 绪论	1
第一节 计算机情报检索的现实意义、作用及其发展状况	1
第二节 国际上几个知名的联机检索系统简介	2
第三节 机检常用术语释义	3
一、一次文献、二次文献、三次文献	3
二、索引及检索工具	3
三、记录、文档、数据库	3
四、顺排档和倒排档	4
五、定题检索与追溯检索	4
六、批处理与联机检索	4
七、字节	4
第二章 国际联机检索技术的基本原理和工作过程	5
第一节 简例：查找有关聚丙烯制备的专利文献	5
第二节 联机检索的原理、步骤和工作过程	7
一、原理简介	7
二、步骤分析	8
三、系统的工作过程	9
四、简图小结	15
第三章 美国 DIALOG 国际联机检索系统	17
第一节 概况	17
第二节 数据库文档简介	18
第三节 常用算符及简单语法符号	18
一、逻辑运算符及其功能	18
二、全文检索符 (Full-text Operators) 及其功能	35
三、截词符“?”及其功能、用法	40
四、其它符号的用法	42
第四节 系统的指令、功能及用法	44
一、系统指令一览表	44
二、指令的一般格式	47
三、常用指令的格式、功能及用法举例	47
第五节 打印格式及检索结果的输出	56
第六节 系统能力的制约	56
第四章 DIALOG 系统化学化工专业数据库的检索	58
第一节 CA 手检历史回顾	58

一、概况	58
二、手检技术回顾	59
第二节 CA Search 系列文档及其记录的构造	61
一、文档特点	61
二、记录的构造	61
三、限定指令的用法	68
四、输出格式代码说明	69
五、可排序字段	73
六、直接提取记录	73
第三节 CA Search 检索用工具书简介	73
第四节 CA Search 系列文档检索实例	74
例 1. 阻燃增强聚丙烯	74
例 2. Cl 化学专利	76
例 3. 压力、流量或流速、温度对蒸馏的影响	77
例 4. 金属簇化合物中的金属-金属键	79
例 5. 氢化肉桂酸	81
例 6. 牻牛儿醇	82
例 7. 噻二唑类	84
例 8. 聚氯乙烯联结剂	85
例 9. 农药 DDT 的残留物	87
例 10. 脾脏的药物代谢作用	89
例 11. Christopher J. Lawson 在 1980—1981 年间的著作	93
例 12. IBM 公司半导体器件专利	94
例 13. 二氯苯	96
第五节 其他化学化工文档结构及用法	97
一、CIN 文档	97
二、Chemsubs 系列文档	106
第六节 总索引 411 文档的特点及用法	123
一、文档特点	123
二、sf 指令	127
三、s 指令	128
四、用法举例	130
五、三个应该注意的问题	136
第七节 串档检索	136
例 1. 利用氨基酸的外消旋作用确定化石年代	137
例 2. 脱沥青油、带料腿结构的分离器	140
例 3. 以尼龙 610 为原料织物的专利	143
例 4. 鉴别三个络合物是否是首次合成	147
第五章 系统数据库的扩充和功能的增强	151
第一节 系统数据库的扩充和调整	151
第二节 系统检索软件功能的增强	152
一、综合部分	152

二、关于全文检索	155
三、关于字段代码的用法	160
四、关于 limit 指令的限定符或参数 (limit qualifiers 或 limit parameters)	166
五、关于检索一定年限内的文档记录	170
六、关于检索一段连续的索引条目	173
七、关于截词技术	176
八、关于联机打印、显示和脱机打印指令	177
第六章 联机检索产品的质量指标	186

第一章 绪 论

第一节 计算机情报检索的现实意义、作用及其发展状况

检索,就是查找的意思。

所谓情报检索,就是要在浩如烟海的大量情报资料中,迅速查找出针对性强、有参考价值的文献、数据等资料。

纵观现代科学技术的任何发明创造或理论成果,不是横向联系国内外科学技术的进展水平,就是纵向联系以往的科学技术成就。而这种联系在很大程度上是靠科技文献资料的交流来建立的。

科研工作者,都深知查阅文献资料的重要性。开题前,要追溯检索,了解与研究课题有关的前人成果;开题后,还要不断收集新发表的文献,了解专题的世界动向。只有在前人成果的基础上来探索未知,科研工作的起点才能立得高,科研工作才能少走弯路。

情报工作者,要写出好的调研报告供领导决策,或向有关方面及时提供情报,除进行实地采访调查外,大多数情况也都是从情报资料的搜集和检索开始的。

经济工作者以及各行各业的高级管理人员,要使自己所从事的工作得到提高,得到较好的经济效益,在技术上、经济上立于不败之地,都必须搜集、分析各种经济及专业情报,他们也处处碰到情报检索问题。

由于科学技术和生产的迅速发展,科技文献和各种经济信息急剧增加。以科技文献为例,近年来每年发表的数量以百万篇计。数量大,类型多,来源广,更新快。有人估计,七十年代以后,其增长速度每5—8年就翻一翻。经济信息,涉及面更广,其品种、数量之繁多,比文献资料有过之而无不及。因此,有人称当今为“信息爆炸”的时代。

我们再来看看情报检索方面的情况。据六十年代美国和日本的调查,某些化学研究人员的业务时间分配比率是:查找阅读资料的时间占51%,实验时间32%,写报告论文时间9%,计划与思考时间8%。采用这个时期的数字是因为当时西方情报工作尚未现代化,检索仍沿用人工手检。事实上,人们面对浩如烟海的文献、数据、经济信息等资料,用传统手检方法来查阅或跟踪,有时50%的时间也不够用,有时甚至是人力所不及。

自1954年美国海军兵器中心首次把电子计算机技术用于情报检索,研制世界上第一个计算机检索系统(以下简称为机检系统)以来,在短短的一二十年中,迫于科学技术和经济发展的需要,利用了科学技术和经济发展的成果,机检技术在全世界得到飞速的发展,经历了以下几个阶段:

1. 五十年代末期,电子计算机技术开始登上图书情报工作舞台;
2. 六十年代赢得了社会承认,正式向公众提供服务;
3. 七十年代与现代通讯技术相结合,联机情报检索技术蓬勃发展,成为社会上“信息产业”的一部分;
4. 八十年代,电子计算机技术正向网络化方向发展。

总之,七十年代国际联机检索的成功及其推广应用,为快速获取全球的科技文献资料提供了渠道,它超越国家和地区的限制,极大地提高了情报的可获得性;因此,它一出现,就赢得了人们广泛的兴趣与重视。

我国的计算机检索事业,虽然开始较晚,进展还是较快的。自1975年起步以来,已经从研究试验阶段,转入到了实用阶段。先后有七十多个单位进行了这方面的试验与服务工作。专门用于情报检索的大、中、小型计算机有十余台,引进的西文文献磁带有卅余种。化工部、石油部、农牧渔业部、地质部、机械工业部等部门的科技情报所,北京文献服务处、上海科技情报所、南京大学等单位,直接或通过归口部门,先后开展了国内的西文情报资料的检索服务,拥有数千个用户。关于中文机检系统的研制,我们国家正在组织制定全国性机检规划,并列为重点项目组织实施。

我国采用国际联机检索技术,由于通讯设备落后,起步更晚一些。1980年中央八个部和一个局联合起来,利用香港的通讯条件,在香港租用了一台情报检索终端机,首辟联机检索渠道,为国内服务。但因为香港与内地的联系走的是传统邮路,信息传递的层次又多,不能发挥联机检索的快速和直接对话的作用,检索时间和检索质量都不能满足要求。81年后,有些中央部属情报所,在北京利用电传终端机(telex terminal)及已有的电报线路,与美国DIALOG系统、ORBIT系统联机,在一定的专业范围内,为用户检索,取得了较好的效果;但邮电部门依外事公文电传收费,通讯费用昂贵(约为香港机检通讯费用的30倍),这样就限制了使用。83年6月,在联合国教科文组织的协助下,中国科学技术情报研究所与邮电部门联合,在北京装设了情报检索用的数据终端机(data terminal),为全国各省市及中央各部举办了应用欧洲ESA机检系统的学习班;同年9月成立了国际联机检索服务部,开展了国际联机检索服务。改建的数据通讯线路,虽然带终端机的能力有限,但通讯费用降低了许多,用户迅速增加。此后,上海、南京、广州、长沙、包头、沈阳、重庆等地,都先后建立了国际联机终端,为地区进行国际联机检索服务。

由于国际联机检索是一种国际间的情报资源共享,其实质是租用技术先进国家的机检系统的分时操作,不需基建投资,即可实现西文资料检索的现代化;所以,一经引入国内,发展的势头就很猛。当然,要利用国际联机检索技术,全面实现我国西文资料检索的现代化,还首先要解决全国各地的远程数据通讯问题,或是降低现行机检电传通讯的收费标准。现在国家对发展邮电事业、发展数据通讯,正在采取正确的政策和措施,如我国研制的通讯卫星已经上了天,并拟改造现行通讯网络等等。总之数据通讯问题得到解决后,随着科研和生产的发展,及对外开放的需要,国际联机检索在全国各地区、各行业、各单位,也将会有更大更快的发展。我们在这样的形势下学习和掌握计算机情报检索技术,特别是国际联机检索技术,是很有现实意义的。

第二节 国际上几个知名的联机检索系统简介

目前国际上较大的联机检索系统有100多个(美国60多个,日本20多个,西欧和加拿大20多个);文档有1100多种。其中有名的如美国Lockheed导弹和空间公司的DIALOG系统,从1972年开始对外服务;最近十多年来,该系统几次更新硬件,不断地改进软件功能,增加数据库文档,改进文档结构,1985年1月已发展成一个拥有二百多个文

档、软硬件功能都很强的情报检索系统；它以数据库包括的内容广泛、文档多、说明书资料等齐备而著称世界。美国系统发展公司的 ORBIT 系统(取 Online Retrieval of Bibliographic Information Timeshared 的缩写而成),文档数量虽不及 DIALOG, 却以独家拥有石油、炼油等文档而著称。欧洲空间组织的 ESA-IRS 系统(European Space Agency-Information Retrieval Service), 据 83 年资料介绍,它的数据库拥有 50 多个文档, 软硬件功能也强,且有独到之处。此外还有美国国家医学图书馆的 MEDLINE 系统, 日本国家科技情报中心的 JOIS 系统及 BRS, NLM, CIS, ADP 等知名系统。

本书将结合美国的 DIALOG 系统, 对国际联机检索技术作较详细的介绍, 以见一般。

第三节 机检常用术语释义

在进入正文之前, 首先对机检常用术语作一些通俗的解释。

一、一次文献、二次文献、三次文献

一次文献即原始文献。如期刊论文、科技报告、学位论文、会议资料、专利说明书等都是。

二次文献是一次文献的浓缩品, 是对一次文献进行分析加工、标引及摘要的记载。标引是依一定的规则, 将一次文献的内外特征, 用非常简洁准确的词语或符号系统, 标示出来。

三次文献是指通过二次文献, 选用一次文献的内容而编写出来的成果; 如专题述评、学科年度总结、动态综述、进展报告、数据手册、百科全书等。

从文献检索来说, 一次文献和三次文献是检索的对象, 而二次文献则是检索的手段与工具。

二、索引及检索工具

索引是以二次文献的某项特征为键, 依特征的字顺或数字大小顺序排列, 后附二次文献序号的集合。如作者索引, 就是以作者为键(作者是文献的外部特征之一), 依作者姓名字顺排列, 后附二次文献序号的集合。又如主题索引, 就是以主题为键(主题是文献的内部特征之一), 依各种主题词的字顺排列、后附二次文献序号的集合。索引是检索工具的重要组成部分, 可引导读者迅速查到所需文献线索。索引的英文词是 index。

检索工具是索引和二次文献等的集合。它将分散的、无组织的二次文献, 依文献的门类及搜集的先后顺序, 组织起来, 给以序号, 这样读者才有可能通过它们的各种索引找到二次文献序号, 从而找到二次文献, 继而找到一次文献。检索工具的重要性之一就在于可以从它迅速查找到一次文献的主要内容及出处。

三、记录、文档、数据库

记录、文档、数据库是机检系统中, 常用到的三个名词。记录是二次文献按机检格式要求的转换物, 其内容与传统二次文献相同; 所以, 一个记录, 相当于一篇二次文献, 对应

于一篇原始一次文献。记录的英文词是 record。

文档是记录及其索引等的集合，是传统的检索工具在计算机检索中的同义物。如美国化学文摘及其有关索引是一套许多年来为各行各业广泛采用的检索工具书，它就相当于机检系统中的一种文档。文档的英文词是 file。

数据库是文档的集合。最小的数据库可由一个文档组成。所以有时文档和数据库同义。数据库的英文词是 database。

四、顺排档和倒排档

顺排档和倒排档也是机检中常用的名词。

顺排档 (linear file) 是二次文献或记录依序号大小排列的集合。传统的检索工具书的正文部分，就是一种顺排文档。

倒排档 (inverted file) 是区别于顺排档的一种文档结构，它以记录的特征为键，按特征字顺或数字大小顺序排列，相当于传统检索工具书的索引部分。倒排档中没有存贮记录的全部内容，仅存贮有文献特征和具有这种特征的记录号。检索者将对文献的需求转化为文献的特征，通过倒排档查到记录号，再通过顺排档查到记录。所以顺排档在计算机检索中，通常是作为辅助资料档，与倒排档并用，二者共同组成一个完整的文档。

五、定题检索与追溯检索

定题检索，简称为 SDI，是 Selective Dissemination of Information 的缩写，相当于代查现刊情报资料。它根据用户所预定的检索内容，定期从新到的文献资料中检出并连续主动向用户提供有关检索内容的最新情报。此项服务适用于研究进程中追踪同类专题的动态和进展。

追溯检索，简称为 RS，是 Retrospective Search 的缩写，相当于普查过刊情报资料。它根据用户的要求，从需求情报时起，追溯到过去某个时间止，是一次性检索，一次提供若干年内的有关情报资料。此项服务适用于开新课题时，全面系统了解有关课题的文献线索。

六、批处理与联机检索

批处理 (batch process)，用户不直接使用计算机，由系统检索人员将用户课题或提问，积累成批进行检索，再分别向用户提供检索结果。目前国内的定题检索服务，一般采用批处理方式。

联机检索 (online search)，用户或专业检索人员，直接使用终端，与机检系统主机相连后，通过人机对话进行检索，并立即获得有关情报资料。联机检索需有终端设备和通讯线路。

七、字节 (byte)

字节也叫“位组”，是作为一个单位来处理的二进制数位，例如 4 位、6 位或 8 位。最常用的是 8 位二进制数。

第二章 国际联机检索技术的基本原理和工作过程

国际联机检索,对不了解情况的人来说,可能是一件神秘的事情,其实并不深奥。我们首先用一个简单的实例,作示意性介绍,以便了解国际联机检索的全过程,然后要再进一步介绍国际联机检索的原理、步骤和工作过程。

第一节 简例: 查找有关聚丙烯制备的专利文献

借助于一台情报检索终端机(一种外形类似打字机的装置,附有打印装置有的还附有显示屏),通过卫星通讯网络,用事先与某机检系统商定的口令(password,是一种密码)呼叫,经系统核定认可后,即可联机检索。检索者在对话中键入系统能理解的各种指令和数据信息,用以指导系统的工作:如在系统的数据库中,查找什么文档;要求命中的文档记录具备哪些信息条件;将命中记录的哪些内容打印输出等等。

下面就是本例与 ESA 系统的联机对话记录的抄件。

```
①      :
②③ ? b2
④ ----- 08 Mar 84 3: 07: 49 User 4564 -----
          0.35 AU  0.25 Minutes in File 10
          0.35 AU approx total

File 02: CHEMABS: VOL66-100,06
(CORP. 1984 BY THE AMER. CHEM. SOC.)
      SET      ITEMS      DESCRIPTION(+ = OR; * = AND; - = NOT)
⑤ ? s rn = 9003 - 07 - 0(w)preparation
      1          202      RN = 9003 - 07 - 0(W)PREPARATION
? s rn = 9003 - 07 - 0(w) manufacture
      2          1109      RN = 9003 - 07 - 0(W)MANUFACTURE
? s rn = 9003 - 07 - 0(W)production
      3          8        RN = 9003 - 07 - 0(W)PRODUCTION
⑥ ? c 1-3/+
      4          1319      1-3/+
⑦ ? 14/pa
      5          1141      4/PA
⑧ ? pr 5/4/1-1141
PRINTED 5/4/1-1141
⑨ ?logoff
----- 08 Mar 84 3: 14: 44 User 4564 -----
          8.07 AU  6.92 Minutes in File 02
          330.89 AU 1141 Offline Prints
```

现逐条解说如下:

① “:”,表示略去的部分,即与系统联机前,用通讯代码及用户密码口令呼叫的部分。其下是全部人机对话。

② “?”是系统的询问符,问检索者要求系统为他做什么事情。“?”出现在检索开始或系统完成了已给出的指令之后。“?”后即是检索者给出的指令。检索者给出的指令均用小写的英文字符表达。如“b2”、“s rn = 9003-07-0(w)preparation”、“c1-3/+”,“pr 5/4/1-1141”等。

③ “b2”是检索者给出的第一条指令。其中“b”是“begin”的缩写,是系统选择数据库文档时,常用的指令之一。“2”是美国化学文摘 ESA 系统的磁带版本 CHEMABS 文档的代号。检索者通过“b2”指令,通知系统在 CHEMABS 文档中查找。

④ 系统将指令的执行结果,紧随指令之后用大写英文字符表达。“b2”指令后的一段,即系统对此指令的回答,告知检索者以下几件事:

(1) 联机时间从 84 年 3 月 8 日 3 时 7 分 49 秒(意大利时间)算起;4564 是用户号;对话前呼叫占用机时 0.25 分钟,占用的是 10 号文档,机时费为 0.35AU(“AU”是欧洲的货币单位,1AU 大约相当于 1 美元。10 号文档即 ESA 系统的 ISMEC 文档,译为机械工程文献题录,调用 10 号文档 1 小时机时费为 84AU,则 0.25 分钟的相应机时费为 0.35 AU)。

(2) 系统已作好查找 2 号即 CHEMABS 文档的准备,此文档从 66 卷开始,已累积更新到 100 卷第 6 期。

(3) “SET ITEMS DESCRIPTION”是一个表头,检索的结果将在这个表头的诸栏下显示:

在 SET 栏下,显示的是记录的集合号,如例中 SET 下的“1”,“2”,“3”,“4”,“5”均为集合号。

在 ITEMS 栏下,显示的是集合中的记录篇数,如例中的“202”,“1109”,“8”,“1319”,“1141”诸数字,分别表示集合 1, 2, 3, 4, 5 中的记录篇数。

在 DESCRIPTION 栏下,将复述检索者在指令中给出的数据或信息,或逻辑组配式,如 rn = 9003 - 07-0(w)preparation 等,但用英文大写字母表达为 RN = 9003 - 07-0(W)PREPARATION 等。

(4) “+”与“OR”相等,同是逻辑或运算符的表达形式;“*”与“AND”相等,都是逻辑与运算符的表达形式;“-”与“NOT”相等,是逻辑非运算符的不同表达形式。联机检索系统一般都采用这三种逻辑运算符。

⑤ 随后一段是检索者给出的第 2—4 条指令及其执行结果。其中“s”是“select”的缩写,是系统常用的检索指令之一;“rn”是“registry number”的缩写,意即登录号;“9003-07-0”是聚丙烯的登录号,代表聚丙烯;括号内的“w”,是一种联接并表达其左右两边数据(或数字字符串)位置关系的符号,w 符要求在文档记录中,右边数据紧跟在左边数据之后,即要求记录中登录号“9003-07-0”之后出现“preparation”,或出现“manufacture”,或出现“production”。数据间的这种关系,就表达了“聚丙烯制备”这个主题。这里,为提高

查全率,检索者采用了表达“制备”的三个英文词,并通过第2—4条指令向系统表达了查找有关聚丙烯制备文献的要求。

用英文大写字母表达的诸行,是系统对第2,3,4条指令的回答,将指令执行结果逐条返回给检索者,即:执行指令2的结果,放在集合1内,有202篇记录;执行指令3的结果,放在集合2内,有1109篇记录;执行指令4的结果,放在集合3内,有8篇记录。

⑥ “c1-3/+”是检索者给出的第5条指令。“c”是“combine”的缩写,是系统的逻辑运算指令,示意系统对集合1,2,3,进行逻辑或运算。系统的回答是对集合1,2,3进行了逻辑或运算后,得集合4,其中有1319篇记录。由此可见,集合1,2,3内的记录,并不重复(因1319为202,1109,8三者之和)。在对话方案中,采用与制备有关的三个英文词来提高文献的查全率,是完全正确的。

⑦ “l”是“limit”的缩写,是系统的限定指令。(请注意数字“1”和英文小写字母“l”的区别)。“l4/pa”意即只要集合4中的专利文献。“pa”是“patent”的缩写。

系统执行上述限定指令的结果,得集合5,内有1141篇专利。

⑧ “pr”是“print”的缩写,是系统的脱机打印指令,其后“5/4/1-1141”告诉系统以格式4(系统的打印格式之一,“4”表示全记录格式,即打印记录的全部内容)、打印集合5内的全部记录,从第1篇开始直至第1141篇。

大写的一行是系统回答已依指令执行完毕。

⑨ “logoff”是脱机指令,通知系统终止检索。

系统回答:

(1) 脱机时间是3时14分44秒;

(2) 以上与2号文档联机对话时间是6.92分钟,机时费为8.07AU;

(3) 脱机打印记录的篇数及费用分别为1141篇;330.89 AU;

(4) 全部费用为338.96 AU。

从简例可以看到,仅花6.92分钟(机时),就检索出1967—1984年2月长达17年CA收集的有关聚丙烯制备专利文献1141篇。因为是脱机打印,约十天后可收到这1141篇专利记录的航空邮寄件。

简例是通过欧洲ESA系统进行的。当然我们也可以通过美国DIALOG系统、ORBIT系统或其它合适的机检系统来查找有关文献。

各联机检索系统的功能,指令的种类、数量及表达形式虽不尽相同,但也都大同小异。我们在简例中初步接触到的ESA系统“b”,“s”,“c”,“l”,“pr”,“logoff”六条指令,是许多机检系统最常用到的。DIALOG系统的全部指令、表达形式及用法,将在本书后面的章节中详述。

第二节 联机检索的原理、步骤和工作过程

一、原理简介

情报检索一般包括存和取两个方面。存,即是组织检索工具。如搜集各种类型的情报资源,制定各种类型的词表,对原始文献进行分析,按一定的规则标引文献的里、表特征,作文摘,选定某些特征为键,依一定的顺序排列,组成索引体系等等。取,即是利用检

索工具,查找所需的文献资料。存是取的基础,取是存的相反过程:只有对已存贮了的情报信息,依存贮时的规则来查找,才能取得出来。

联机检索系统如何实现情报检索的呢?人工查找资料,多是用手翻动工具书、用眼脑扫描各种索引、题目等,与脑中已明确的对文献的需求相比较,来找到自己所需资料并摘抄下来。计算机检索是对人工检索智能的模仿。但计算机不认识书本上的文字,不能读书,所以必须首先把题目、文摘、索引等检索工具书本上的文字信息、转换成计算机能识别的由“0”,“1”两个字符不同排列组成的代码,并记录在磁带或磁盘上,这才构成了计算机可读文献,即前面谈到的数据库文档。有了文档,还要有内外存容量大、运算速度快的计算机硬件和逻辑思维能力强的检索软件。硬件、软件、数据库三者齐备的系统,为检索提供了物质条件。但这样一个系统还不知道人们的要求和人们查找文献的思路,即人们需要什么文献和怎样找到这些文献。人们对文献的需求及查找文献的方法(可转换为所需文献的特征及查找文献的途径等),要在系统的功能及指令系列的制约下,用系统能理解的语句,从终端机上键入;再通过国际卫星通讯网络传达给系统,系统才能理解检索者的意图,并据此进行工作。系统的检索结果,也是通过国际卫星通讯网络传送,在终端机上输出告诉检索者。

二、步骤分析

在理解了联机检索原理以后,对照简例,可以把联机检索全过程分为以下几个步骤。

(1) 对要检索的课题进行分析,明确情报需求。

(2) 选定一个与你有关联(即你已取得了该机检索系统的 password)并为你所熟悉的国际联机检索系统与之联通。

(3) 选定要查找的文档。

(4) 用系统能理解的语言来表达你对情报的需求和查找情报的方法,系统据此进行工作,直到获得满意的检索结果。如同手检工具书多用英语一样,大多数国际联机情报检索系统的语言(包括指令及数据库数据等),也采用英语。所以,对我们中国人来说,还必须将表达情报需求的检索用词,由汉语形式转换为英语形式。

(5) 命令系统将检索结果依一定格式打印输出。

(6) 脱机。由于机检是以电子的速度进行、是以分秒计费的,若是查找失误或一时得不到满意的结果,应该立即与系统脱离。得到了满意的检索结果,也应该及时发出脱机的终结指令,系统此时将告诉你检索费用等。

检索费用分数据库文档机时租金、通讯费、资料打印费三部分。机时租金随查找的机检系统及文档的不同而不同。以 DIALOG 系统为例,最便宜的如 1 号教育文摘,25 美元/时;最贵的如 223,224,225 号化学专利单元词索引,高达 300 美元/时。通讯费,电传依现行价格每分钟人民币 7.2 元,采用数据通讯方式则便宜一些。资料打印费,依不同文档的不同格式,价格均不一样。一般最简格式约数美分/篇,最详格式为数十美分/篇;个别文档如 220,221 号美国专利引文索引,129 号三十五国市场预测报告的全记录打印费,则高达 50 美元/篇以上。所以,一个专题的检索费用,不是固定不变的,它随查找的机检系统的不同而不同,随查找的文档及文档数量的多少而不同,随检索途径的多少、检索的深度、广度而不同,随打印的资料篇数及格式而不同。检索者检索技术水平的高低,操作的

熟练程度,对费用都有很大的影响。一个题目的检索费用,从几十美元到数百美元不等。当然检索费用也可以根据文献需求者能付得起的金额由检索者进行调节。

以上六个步骤也可以概括为: 上机前的准备和联机对话操作两大步。

三、系统的工作过程

在联机检索原理简介中曾提到,系统是根据检索者的指令进行工作的;系统是硬件、软件、数据库三者齐备,且与国际卫星通讯网络相连,后者能快速远距离传输检索者与系统相互所需要的指令、信息、数据等。下面介绍系统是怎样处理所得到的指令及信息数据,并将检索者所需文献检索出来的。

(1) 在回答上述问题前,首先要了解机检系统数据库的结构。

机检系统的数据库,是一个庞大的机用二次文献图书馆。这个图书馆保存的不是书本,而是载有二次文献及其索引的磁带或磁盘。由这些磁带或磁盘构成数据库文档。一个机用系统的数据库,一般拥有几十到数百个文档。每种文档通常又由一个顺排文档和二一个倒排文档构成。顺排档是文档全部记录依系统存取号大小顺序排列的集合;存取号愈大,记录愈新。二个倒排档,一个是基本索引(basic index),一个是辅助索引(additional index)。基本索引是文献内部特征的集锦;它搜集了记录上题目(TITLE,取TI为其标识符,下同)、主题词(DESCRIPTION——DE)、自由词或关键词(IDENTIFIER——ID)、文摘(ABSTRACT——AB)等字段(这些字段从不同侧面描述了文献的内在特征)的单词及词组作为它的条目,依英文字母排序,后附记录的存取号。辅助索引是文献外部特征的集锦;它搜集了记录上AU(AUTHOR,作者)、JN(JOURNAL NAME,期刊名称)、LA(LANGUAGE,语种)、PY(PUBLICATION YEAR,出版年代)等字段(这些字段从不同侧面描述了文献的外部特征),先依字段标识符字母排序,然后依字段内容字母或数字大小排序,后附记录的存取号。

图 2.1、图 2.2、图 2.3 为某一数据库文档以二篇记录为例的顺排档和倒排档的片段。

从图 2.1 可以看到:记录在顺排档中依存取号排序;记录由不同的字段组成,每个字段都有自己的标识符。对基本索引诸字段的词,还标出了在字段中的位置顺序,如 TI1, TI2, TI3, ..., AB1, AB2, AB3, ...; DE1, DE2, DE3, ...。

图 2.2、图 2.3 为与图 2.1 记录相对应的在倒排档中的片段。图 2.2 是基本索引片段,图 2.3 是辅助索引片段。从图 2.2 可以看到:基本索引的条目,由描述文献内部特征诸字段(此处是 TI, AB, DE 三个字段)的词及词组组成,依字母排序,条目后附记录号、字段标识符及位置顺序号,表明条目的来源或出处。从图 2.3 可以看到:辅助索引的条目,由描述文献外部特征诸字段组成,首先依字段标识符字母排序,然后依“=”号后字段内容的字母或数字大小排序,后附记录号,表明了条目的来源或出处。必须特别指出的是,在辅助索引中,字段标识符和等号是作为条目前缀出现的;而在基本索引中,字段及位置标识符则出现在存取号之后。在后面还将看到,基本索引诸字段的字段标识符,若检索者将它们作为所需文献特征条件之一对检索项加以限定时,将以后缀形式附在词尾,如 teaching/DE, communication/AB 等。

从以上描述不难理解,倒排档的实质,是一种文献内、外特征出处的调查统计一览表;可由此计数出具有某特征的记录在文档中出现的频率或拥有的记录数,以及这些记录的

存取号。

联机检索系统有 9 个禁用词 (stop words) 列于图 2.4 中。这些禁用词因出现频繁, 且无实词意义, 为节约内外存, 不列入基本索引中。但为配合位置关系符 w 等的运用, 在图 2.1 记录上仍标明了它出现的字段位置。

EJ218653 (accession number) Nonverbal Communication in Teaching TI1 TI2 TI3 TI4 Smith, Howard A. AU Review of Educational Research, v49 n4 p631-72 Fall 1979 JN PY Language: English LA School-related research is reviewed under seven categories of nonverbal AB1 AB2 AB3 AB4 AB5 AB6 AB7 AB8 AB9 AB10 communication: environmental factors, proxemics, kinesics, touching AB11 AB12 AB13 AB14 AB15 AB16 behavior, physical characteristics, paralanguage, and artifacts. AB17 AB18 AB19 AB20 AB21 AB22 Descriptors: *Classroom Communication/ *Nonverbal Communication/ *Teaching DE1 DE2 DE3 DE4 DE5
EJ218652

EJ195895 (accession number) Understanding Body Language TI1 TI2 TI3 Sielski, Lester M. AU Personnel and Guidance Journal, v57 n5.p238-42 Jan 1979 JN PY Language: English LA The article details the significance of nonverbal communication or AB1 AB2 AB3 AB4 AB5 AB6 AB7 AB8 AB9 body language of the counselor in interview situations. AB10 AB11 AB12 AB13 AB14 AB15 AB16 AB17 Descriptors: *Body Language/ *Counselor Role/ *Helping DE1 DE2 DE3 DE4 DE5 Relationships/ *Nonverbal Communication DE6 DE7 DE8
EJ195894

图 2.1