

科技文献工作培训教材

国际联机检索

韩玉宝

湖南省科技情报研究所

一九八六年九月

前 言

国际联机检索是一项产生于六十年代，在七十年代得到迅速发展的现代情报技术。

第二次世界大战后，科学技术迅猛发展，科技文献量也以相应的速度迅速增长，作为科学研究活动重要组成部分的情报检索，以传统的手工检索方式，在飞速增长的文献量面前，变得愈来愈繁难，愈来愈费时，这一由科学技术的发展而带来的结果，反过来严重地影响了科学技术的发展。

六十年代中期，在计算机技术和通讯技术发展的基础上，国外首次研制出了联机检索技术。此后随着同步通信卫星技术的完善，使联机检索超出了一国的范围而成为今天的国际联机检索技术。

国际联机检索的实现，使世界各地的情报需求者能够越过地理空间的障碍，在瞬息之间获得所需要的情报资料，共享全世界的情报资源，这是对传统的手工检索情报资料的一场深刻的革命。

国际联机检索的历史及现状

目 录

前 言	国际联机检索的前身是联机检索，即使用终端设备从输入机里的数据库中检索情报资料，最早要追溯到五十年
一、国际联机检索的历史及现状	(1)
二、国际联机检索系统的组成	2)
三、国际联机检索的工作形态	(4)
四、联机检索的工作语言	(5)
五、国际联机检索的基本方法	(6)
六、联机检索的准备和上机	(44)
参考文献	(59)

目前国际联机检索在世界上已相当发达，全世界已有各种数据库1700多个，全世界90%的文献资料进行联机检索，检索终端已近百万台。

具有检索功能的联机系统已达100多个，其中规模最大的是美国洛克希德公司的DIALOG系统，美国系统发展公司的ORBIT系统，欧洲空间组织的ESA/ISIS系统，美国国家医学图书馆的MEDLINE系统，日本国家科技情报中心的JQS系统。

由于国际联机检索在世界范围内的发展，我国在八十年代开始发展国际联机检索，中国科学技术情报所、中图一些

一 国际联机检索的历史及现状

国际联机检索的前身是联机检索，即使用终端设备从输入到计算机里的数据库中检索情报资料，最早要追溯到五十年代。1954年美国海军军械中心（NOTS）使用IBM—701型计算机建立了世界上第一个计算机情报检索系统，检索者可以近距离地使用终端从这个系统检索情报资料。

随着通信技术和通信网络的发展，终端设备可以愈来愈远离数据库中心。特别是由于六十年代卫星通讯技术的发展，检索者可以从地球上的任何地方从数据库中进行联机检索，产生了真正的国际联机检索。

国际联机检索在七十年代得到了迅速的发展，已成为一项成熟的现代情报技术，这项技术对传统的手工检索是一场深刻的革命。

目前国际联机检索在世界上已相当发达，全世界已经有各种数据库1100多个，全世界80%的文献可以进行联机检索，检索终端已达几百万台。

具有对话功能的联机系统已达100多个，其中规模较大的有美国洛克希德公司的DIALOG系统，美国系统发展公司的ORBIT系统，欧洲空间组织的ESA/IRS系统，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系统等。

由于四化建设的需要和对外交流活动的发展，我国在八十年代开始发展国际联机检索，中国科学技术情报所，中央一些

部，一些省级情报所先后建立了国际联机检索终端，目前已达四十几台，在短短的几年里，国际联机检索已经显示了强大的生命力和优越性，可以预见，国际联机检索将对我国的四化建设作出巨大的贡献。

二 国际联机检索系统的组成

联机检索系统是由主计算机、数据库、通信网络和终端等组成的。

1. 主计算机

主计算机是执行检索功能的主体设备，它根据由终端输入的指令，在数据库中查找符合指令要求的文献，并根据指令的要求对文献进行处理，然后将检索结果传递给终端。主计算机的量级往往取决于数据库的大小。在DIALOG系统中，使用NAS—9000型和IBM—3033型大型电子计算机各一台作为系统的主计算机，单机运算能力达14MIPS（即每秒钟可处理1400万条指令）ESA系统使用SIEMENS7865双机，运算速度为260万次/秒。大型商业性国际联机检索系统中心往往采取双机制，以保证系统工作的高度可靠性。

2. 数据库

数据库是主计算机执行检索的对象，其载体为磁盘或磁带。软件人员对文献进行标引和文摘等前处理后，存入磁盘或磁带，构成数据库。目前，越来越多的文献检索工具刊物出版机构在出版书本型刊物的同时，也出版计算机阅读（简称机读）型检索工具，如《世界专利索引》、《化学文摘》、《工程索引》

等, 均有机读磁盘或磁带出版, 系统数据库稍加处理即可直接使用。

数据库由若干个文档构成, 如DIALOG数据库到目前为止, 有280多个文档, ORBIT数据库有80多个文档, ESA—IRS数据库有70多个文档。每个文档收集某个或某些领域的文献。

在每个文档中, 都包含有数量庞大的记录, 每篇文献经过处理后, 存入数据库的相应文档中, 即构成一条记录。

每条记录由若干字段构成, 字段有两种类型, 一种是基本字段, 一种是辅助字段。基本字段反映文献的内容特征, 如标题字段、文摘字段, 规范词字段等; 辅助字段反映文献的外部特征, 如作者字段, 文献发表时间字段、语种字段、刊登文献的刊名字段等。

数据库中的各个文档均定期进行更新, 补充最新的内容, 以保证联机检索的新颖性。

3. 终端

在联机检索中, 终端是检索者用来输入检索指令, 并向检索者显示和打印检索结果的装置。检索者使用终端与主计算机进行人机对话, 实现联机检索

用于国际联机检索的终端, 主要有三种形式: 电传终端、标准终端和微机终端。

电传终端

使用用户电报机作为终端。由于设备本身和电报电路的限制, 有传输速度低(50bps, 即每秒50波特)的缺点, 检索占机时间较长, 检索费用高。

标准终端

是进行联机检索的专用设备, 由键盘、显示屏幕和打印机

构成，传输电路使用电话电路，常用传输速度为300bps和1200bps

微机终端

在微型计算机上加上异步通信支持软件，就可以把微机作为检索终端使用了。

与标准终端比较，微机具有存储和处理功能，可以将检索策略在微机上进行预处理，可免去联机对话时人工键入检索指令，能大大缩短机时。

4. 通信网络

通信网络是联结终端和主计算机的手段，没有通信网络，检索者就不能在千里之外用终端与主计算机对话了。

在国际联机检索中，整个通信网络由有线通信和无线通信两部分组成。无线通信由同步通信卫星和卫星地面站实现，或者由地面微波中继站实现。

三 国际联机检索的工作形态

联机检索时，检索者首先将终端置于联机状态，根据规定的通信程序，在终端上输入网络密码和系统主机密码，接通终端与主机之间的通信线路。

线路接通后，检索者就可以进行检索了。检索者根据预先编制好的检索策略，逐条向终端输入检索指令，终端将这些指令转换为一系列二进制直流脉冲信号，这些信号经过调制解调器，转换为能由通信网络传递的交流信号，交流信号通过有线线路或无线线路（报路，话路或数传线路）传递到卫星地面

站，再由地面站以无线通信方式发送到通讯卫星，通讯卫星将信号重新调制并放大后，发送到与主计算机联结的卫星地面站，地面站将接收到的交流信号经过有线通信传递给主计算机前的调制解调器，调制解调器将交流信号还原成为主计算机能执行处理的二进制直流脉冲信号。

主计算机执行指令的结果以同样的方式由主计算机传递给终端，所不同的只是调制解调器执行着相反的工作。

通讯线路的联结如下面的框图所示：



终端和主计算机之间的通讯采用半双工制，即一方处于发送状态时，另一方只能处于接收状态，不得发送，双方交替地处于发送和接收状态。

四 联机检索的工作语言

检索者的检索思想是通过语言来描述的，联机检索的工作语言是终端指令语言，终端指令语言是一种非程序的人机对话语言。主计算机每执行一条由终端输入的指令后，都要将执行的结果传递给终端，检索者根据上条指令的执行结果输入下一条指令，直到完成整个检索。

每个联机检索系统都规定有各自不同的终端指令语言，检索者要检索某一个联机检索系统，就必须严格遵守这个系统的语言语法来编写检索策略。在后面的文章中，我们均使用，DIALOG系统的终端指令语言讲述联机检索的基本方法。

五 国际联机检索的基本方法

本章以DIALOG的系统功能和语言语法为基础，叙述国际联机检索的基本方法。

1. 预备概念

①DIALOG系统的业务项目

- a. 联机检索，是DIALOG系统最主要的功能和业务项目；
- b. 联机打印，以联机状态将检索的结果通过终端输出打印给检索者；
- c. 脱机打印，以批处理方式将检索的结果通过系统的高速打印机打印出来，邮寄给检索者；
- d. 联机SDI服务，检索者通过联机将检索策略存入主计算机中，每当文档更新时，主机自动用已存入的检索策略查找新装入的记录，然后将检索结果邮寄给检索者；
- e. 联机订购原文，以联机方式向系统订购所检结果的原文资料。

②DIALOG的输出打印格式

DIALOG的输出打印格式共有八种：

格式1 DIALOG存取号；

格式2 除文摘外的全记录；

格式 3 DIALOG 存取号、篇名、作者、文献出处和语种；

格式 4 DIALOG 存取号、篇名和文摘；

格式 5 全记录 (DIALOG 存取号、文摘号、篇名、作者、文献出处、语种、文摘、叙词和自由标引词)；

格式 6 DIALOG 存取号和篇名；

格式 7 DIALOG 存取号、文摘号、篇名、作者、文献出处、语种和文摘；

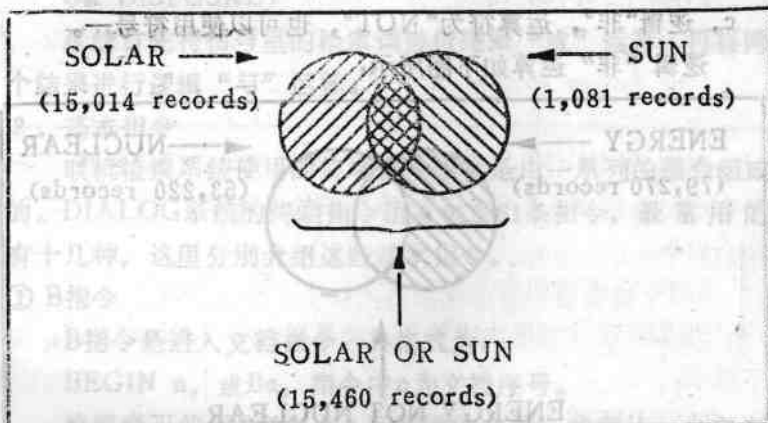
格式 8 DIALOG 存取号、篇名、叙词和自由标引词。

③布尔逻辑运算

为了实现概念和集 (后面讲到) 之间不同的组配关系, DIALOG 系统使用布尔逻辑运算, 有三种布尔逻辑运算符, 分别叙述如下:

a. 逻辑“或”, 运算符为“OR”, 也可以使用符号+。

逻辑“或”运算如下图所示:

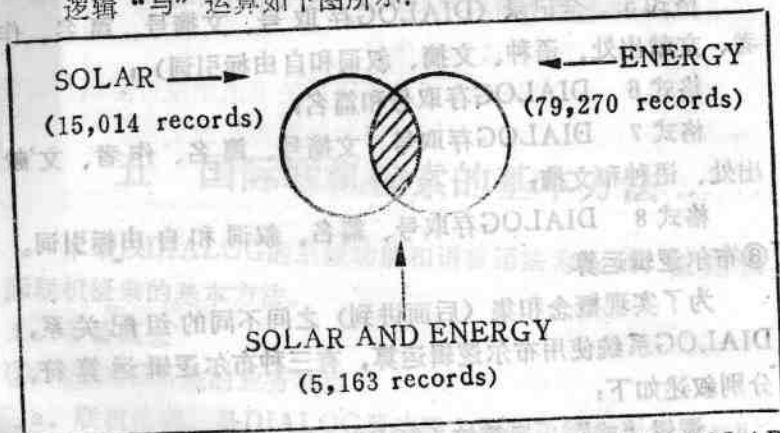


在该逻辑“或”运算中, 系统给出含检索词SOLAR 或含SUN, 以及同时含SOLAR和SUN的文献。

图中的数字为检索词的文献命中篇数。

b. 逻辑“与”，运算符为“AND”，也可以使用符号*。

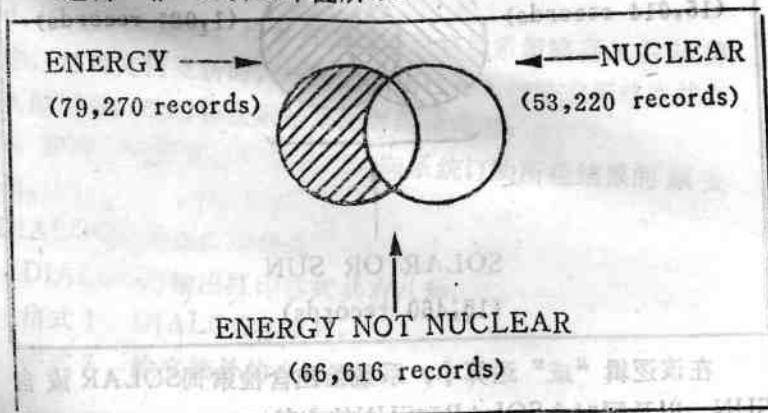
逻辑“与”运算如下图所示：



在该逻辑“与”运算中，系统给出同时包括检索词SOLAR和ENERGY的文献。

c. 逻辑“非”，运算符为“NOT”，也可以使用符号—。

逻辑“非”运算如下图所示：



在该逻辑“非”运算中,系统给出只包含检索词ENERGY,而不含NUCLEAR的文献。

使用逻辑“非”运算时要谨慎,如检索者希望检索关于无氰电镀的文献,将检索词进行如下逻辑运算:

ELECTROPLATING NOT CYANIDE

结果正好把关于无氰电镀的文献去掉了,因为这样的文献中包含了检索词CYANIDE。

在一个检索式里,可以使用两个以上逻辑运算符,系统在执行有两个以上运算符的逻辑运算时,首先执行逻辑“非”运算,其次执行逻辑“与”运算,最后执行逻辑“或”运算。如果要改变系统的执行顺序,可以使用括号,则系统优先执行括号里的逻辑运算,括号外面的逻辑运算则仍按系统顺序执行。

如:

(RADIOACTIVE OR NUCLEAR) AND (WASTES OR DISPOSAL)

系统首先将括号里的检索词进行逻辑“或”运算,再将两个结果进行逻辑“与”运算。

2. 基本指令

联机检索系统使用的终端指令语言是由一系列的指令组成的。DIALOG系统的终端指令语言包含61条指令,最常用的有十几种,这里分别介绍这些基本指令。

① B指令

B指令是进入文档指令,其格式为:

BEGIN n, 或Bn, 指令中n为文档序号。

检索者可使用B指令进入所需要的文档,或者从一个文档转入另一个文档。

下图给出B指令的使用及系统对B指令的响应:

<i>Estimated cost of previous Search</i>	<i>Date</i>	<i>Clock Time (U.S. Eastern Time)</i>	<i>Customer or User Number</i>
? BEGIN 8	17 sep 79	11:21:01	User 3462
\$ 0.39	0.006 Hrs	File 8	
File 8; COMPENDEX 70-79/AUG (Copr. Engineering Index Inc.)			File Banner
<u>Set Items Description (+ = OR</u> <u>★ = AND; - = NOT</u>			
<i>Coverage and Update Status</i>	<i>Default File n or file Used on Previous Search</i>		
<i>Time Elapsed in Previous File Searched or Logon Procedure</i>			

图中“BEGIN”前的问号“?”是系统给出的催促符,检索者在系统给出“?”后,即可向系统输入指令。

② S指令

S指令是选词指令,其形式为:

SELECT 后接检索单词,或简写为S 接检索单词,如下所示:

? SELECT PETROLEUM 或

? S PETROLEUM

S 指令也可以后接用逻辑运算符组合起来的两个以上的

单词或其他数据元素。

? S PETROLEUM AND PRICES AND OPEC
AND PY = 1979

2711 PETROLEUM

2683 PRICES

151 OPEC

101655 PY = 1979

19 83 PETROLEUM AND
PRICES AND OPEC
AND PY = 1979

例中检索词前面的数字为该词命中的文献篇数。19是一个集号，由系统自动按顺序给出。

S 指令还可以直接使用由前面指令响应得到的集号，如：

? S S3 AND COMPUTER

该指令中，3是集号，前面必须加S，S和3中间不留空格。

每一条S指令可以得到一个集号，如果几个检索单词由逻辑运算符组合起来，则系统给出每个检索单词的命中数，前面没有集号，系数只对整个逻辑组合给出集号，如前例。

在S指令中，有如下一些单词不能使用，称作禁用词，这些词是：

AN FOR THE

AND FROM TO

BY OF WITH

如果在S指令中必须使用上述单词,则必须用引号括起来,

如:

? S "OR" (W)GATE

③ SS指令

SS指令是分步选词指令,其形式为:

? SELECT STEPS 后接检索词,或

? S STEPS 或

? SS

SS指令具有S指令的全部功能,遵守与S指令完全相同的语法和规定。不同的是系统响应SS指令时,对每一个检索单元(不一定是单词,可以由位置符联结起来的两个以上的单词,后面讲到)都给出一个集号。

系统的这种响应方式大大地方便了检索者的使用,检索者在后面的指令中,可以任意使用系统给出的集号,来代表相应的检索单元。

? SS PETROLEUM AND OPEC AND PRICES

10 2711 PETROLEUM

11 151 OPEC

12 2683 PRICES

13 83 10 AND 11 AND 12

例中10、11、12,均为系统分配给每个检索单元的集号,13是整个检索式的集号。

④ C指令

C指令是组合指令,指令系统进行逻辑运算,其形式为:

? COMBINE 或

? C

在C指令中, 只能使用集号, 不能使用检索词或其他检索元素。

COMBINE 1 AND 2 = C1 AND2 = C1 AND2 = C1 AND2
COMBINE 3 OR 4 = C3 OR4 = C3 OR 4 = C3 OR 4
COMBINE 5 NOT 6 = C5 NOT6 = C5 NOT6 = C5 NOT 6

C指令可以对两个以上的集号进行逻辑组合:

COMBINE1 AND (2 OR 3)
COMBINE (4 OR 5) AND (6 OR 7)
COMBINE ((8 OR 9 OR 10) AND (11 OR 12))
NOT 13

两个以上的连续集号的“与”或者“或”, 可以缩写为以下形式:

COMBINE 1-3/AND = C 1 AND 2 AND 3
COMBINE 1-3/OR = C 1 OR 2 OR 3

系统在执行C指令后, 对执行结果也给出一个集号。

⑤T指令

T指令是联机打印指令, 指令系统在联机状态下打印检索

结果，其形式为：

? TYPE n/f/r 或

? T n/f/r

指令中n为要求打印记录所在的集号，f为打印格式，r为要求打印记录的条数范围。

如 T 5/6/1—10

就是指令系统以第6种打印格式联机打印集号5中的第1到第10条记录。

系统在输出打印记录时，自动按记录存入系统的时间为序，以最新输入系统的记录作为第一条，依次向前排序。

在r项中，也可以打断顺序任意指定，如 T5/6/3，即要求打印集号5中的第3篇；在T5/6/1-2, 7, 15—16, 10中要求打印的条数范围为几段，中间以逗号隔开。

检索者如果不输入r项，则系统自动按打印指定集号、指定格式的第1条记录处理。之后若只输入T，则系统打印该集号的第2条记录。

检索者如果不输入f项，则系统按第2种格式处理。

⑥PR指令

PR指令是脱机打印指令，系统主机接收该指令后，命令系统中心高速打印机打印检索结果，然后以一级邮政邮寄给检索者。

PR指令的形式为

? PRINT n/f/r 或

? PR n/f/r

n, f和r的含义与在T指令中完全相同，在r项要求条数范围中，系统按时间顺序给出，以最新装入的记录为第1篇，r项可以是一条记录，可以是一个范围，也可以是一串不连续的记