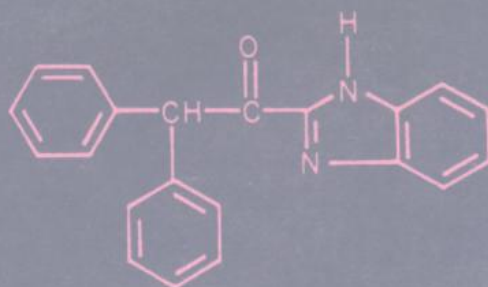
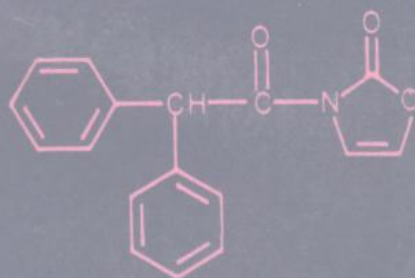


化学物质结构图形 联机检索入门

谢湘音 编著



科学技术文献出版社

化学物质结构图形 联机检索入门

谢湘音 编著

L60767
2007/09

科学技术文献出版社

(京) 新登字130号

内 容 简 介

本书主要介绍国际联机检索在化学化工、药学等领域中的应用,特别是在检索化学物质方面的应用。内容涉及STN国际联机检索系统中,美国化学文摘社(CAS)编制的“化学物质登记数据库”所提供的化学物质检索功能及其方法。此外,还可以利用该检索方法准确、全面的进行化学物质的查新和申报专利的文献检索。本书介绍的结构图形检索方法适用于各种类型的终端,对软硬件无特别要求。

本书适合从事情报检索的人员,以及化学化工、药学等领域各行业的科技工作者,并可作大专院校有关专业参考书。

图书在版编目(CIP)数据

化学物质结构图形联机检索入门/谢湘音编著. —北京: 科学技术文献出版社, 1994. 7

ISBN 7-5023-2121-7

I. 化…

II. 谢…

III. ①机器检索-联机系统-应用-化学工业 ②化学工业-应用-机器检索-联机系统

IV. G354.4+TQ-6

科学技术文献出版社出版

(北京复兴路15号 邮政编码100038)

北京怀柔燕东印刷厂印刷 新华书店北京发行所发行

1994年7月第1版 1994年7月第1次印刷

开本: 850×1168毫米 1/32 11.5印张 294千字

科技新书目: 306—100 印数: 1—2500册

定价: 12.80元

前 言

本书主要介绍当前国际联机检索最新技术之一——图形检索在化学化工、药学等领域中的应用，特别是在检索化学物质方面的应用。其内容涉及STN国际联机检索系统中美国化学文摘社（CAS）编制的“化学物质登记数据库”所提供的化学物质图形检索功能及其方法。书中重点介绍了在与STN国际联机检索系统联机的ASCII终端或微机文字终端上，建立化学物质结构图形及检索图形的方法，并以此为手段获取化学物质情报及其相关参考文献。

目前，随着计算机技术的飞速发展，情报检索技术已向图形检索和数值检索的深度发展，并逐步进入普及阶段。但是，国内对图形检索技术尚未进行推广，除受到检索费用的影响外，主要还是对联机图形检索技术尚不了解和不能熟练掌握其检索技术等原因所造成的。故本书出版的目的是：1. 引进和介绍联机图形检索技术，使之应用于化学化工等情报领域。2. 普及和推广化学情报联机检索技术，使其更好地为广大科技人员服务，为生产建设服务。众所周知，动植物拉丁名称是世界生物学家的通用语言，而对化学家来说，化学物质结构图就是他们的共同语言。化学物质结构图形的检索，解决了世界化学家检索化学情报的语言障碍。也就是说，他们不必再为各种复杂的化学物质名称而困惑，只需利用图形就可以快速、准确地获取化学物质的情报及相关参考文献。同时，由于图形的直观作用，也为非化学专业的情报检索人员打开了检索复杂结构化学物质的大门。3. 解决目前手工检索化学物质的一些困难，甚至一般联机检索都难于实现的化学物质子结构检索技术。为化学物质衍生物、同类物等基础研究及系列产品研究提供了先进方便的手段。

本书共分五章。第一章主要介绍国际联机检索基本知识，其目的是为了读者初步了解STN国际联机系统的基本检索方法。第二

章概括介绍STN国际联机检索系统中美国化学文摘社提供的“化学物质登记数据库”的内容及其检索功能，特别是把结构图作为检索项进行检索的功能揭示给读者。第三章为学习本书重点部分（第四章、第五章）提供一般性基本知识。第四章重点介绍如何在ASCII终端，通过STN系统，利用化学物质登记数据库提供的各种作图指令建立化学物质结构图形。第五章重点介绍如何检索建立好的结构图形，并以此为手段获取化学物质情报及其相关参考文献。

本书参考了美国化学文摘社（CAS）出版的《USING CAS ONLINE——REGISTRY FILE》一书中的第II卷A和第III卷。由于实验经费有限，本书第四章很多实例和第五章的部分实例选用了该参考文献的例子，凡不清楚之处可参见原版书。书内提及的指令和算符的详细说明和使用方法可参见STN系统出版的《Searching STN: The Basics》及《GUIDE TO COMMANDS》（简称“STN用户指南”）。

本书在撰写过程中，得到了不少同志的大力支持，也得到本部门领导的支持，在此谨向张曼华，陈家显等同志及支持本书出版的领导表示感谢。本书插图由韩进泉同志完成。

由于水平有限，书中难免有不妥和错误之处，敬请读者批评指正。

作者

1992年10月

目 次

前言

第一章 国际联机情报检索基本知识	(1)
第一节 与计算机系统联机.....	(1)
第二节 国际联机情报检索基本方法.....	(13)
第三节 STN国际联机系统基本检索技术.....	(18)
第二章 美国化学文摘社“化学物质登记数据库”简介	(29)
第一节 CAS ONLINE 部分数据库简介.....	(29)
第二节 化学物质登记数据库简介.....	(31)
第三章 结构图形检索基本知识	(48)
第一节 STN系统建立结构图形的四种方法.....	(48)
第二节 建立图形使用的终端类型.....	(51)
第三节 概念和名词解释.....	(53)
第四章 ASCII码终端建立结构图形键盘指令方法	(68)
第一节 REG 文档建立结构图形的基本操作过程	(68)
第二节 建立图形骨架——GRAPH(或GRA)指令 和MOVE(或MOV)指令及 RECALL (或REC)指令的应用.....	(74)
第三节 改变节点原子——NODE(或NOD)及 VARIABLE(或VAR)及 REPEATING (或REP)指令的应用	(124)
第四节 确立化学物质结构的键值、键类型—— BOND(或BON)指令.....	(166)
第五节 结构特征属性指令.....	(188)
第六节 其它指令.....	(218)
第七节 化学物质结构图形建立实例简介.....	(247)

第五章	STN 系统结构图形检索的方法	(252)
第一节	REG文档结构图形检索的多种途径 及其组合检索功能	(252)
第二节	用SEARCH(或S) 指令检索化学物质结构图	(253)
第三节	REG 文档结构图形检索的三种“检索范围”	(257)
第四节	REG 文档图形检索的三种“检索类型”	(271)
第五节	化学物质结构图形检索实例	(287)
第六节	结构图形联机检索的一些策略问题	(292)
第七节	REG文档的结构图形更新检索	(314)
第八节	REG文档的其他指令	(323)
附录1	片断文档——系统提供的“结构模型”	(345)
附录2	化学物质结构图形检索实例	(354)

第一章 国际联机情报检索 基本知识

本章主要介绍联机情报检索的基本知识和国际联机情报检索的概念。作为一个要查找科技情报文献的用户,如何通过联机检索,通过向计算机输入检索词(作为检索项),向联机系统索取与该词、该主题有关的情报及相关文献记录。同时,介绍目前世界第二个大型国际联机系统——STN系统的基本检索方法,从而为理解和掌握本书重点内容,即STN系统化学物质登记数据库(REGISTRY 文档或REG文档)提供的结构图形检索方法奠定基础。

第一节 与计算机系统联机

与计算机系统联机的目的就是进行联机情报检索。联机情报检索英文称为 ONLINE INFORMATION RETRIEVAL,也可称为联机检索(ONLINE SEARCHING)。简言之,联机情报检索就是通过自己的终端与大型计算机检索系统联机,实现在自己的终端上与计算机系统“对话”,从而获取计算机系统所拥有的数据库中储存的各种情报文献。而实现联机检索必须具备以下三个要素:1. 终端设备。用以与大型计算机系统连通,作为“人机对话”的工具。2. 通讯线路及设备。通讯线路是实现终端机与大型计算机系统之间信息转换和传递的桥梁。3. 大型计算机系统(也称主机系统,国外称为HOST)。主机系统拥有大量机读文献数据库和储存了各种类型的数据和情报,并且提供获得这些情报的手段(检索软件)。下面用图1.1所示的方法说明与计算机系统联机的全过程。

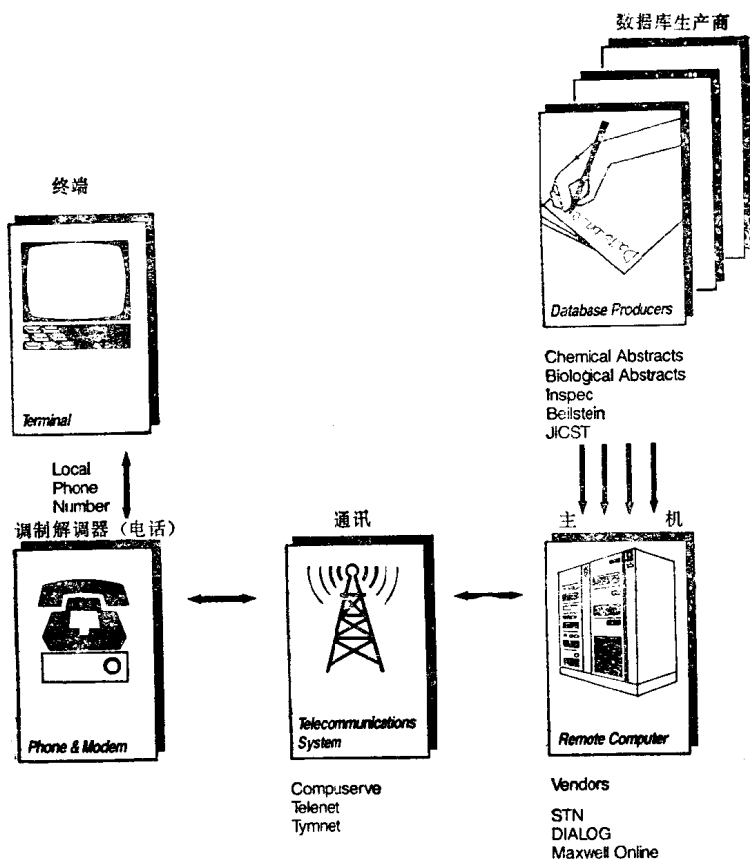


图 1.1 联机检索设置示意图

一、终端设备

终端设备主要由终端机，当然也可以用微机作终端机(包括键盘)，与终端机相连的打印机及调制解调器三部分组成，如图 1.2 所示。

终端机是检索人员用来和主机系统对话的工具(也可用微机代替)，用户向主机系统提出的问题是通过对终端上的键盘输入指令和

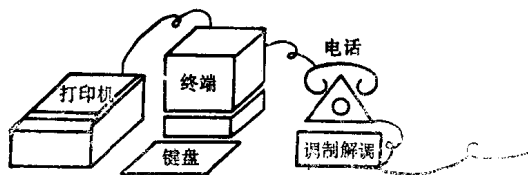


图 1.2 终端设备示意图

参量(在终端屏幕上显示出来), 并使用回车键将指令传送给计算机处理后送往通讯设备、线路, 即首先由调制解调器接收计算机处理完的信号转化成通讯信号, 送往通讯线路, 最后进入主机系统。而主机系统的回答则通过相反的过程和路径最后在终端屏幕上显示出来。这样一来一去、一问一答反复多次逐步完成全部“对话”, 实现检索目的。

打印机的作用是将检索结果及检索人员需要的情报打印出来。

调制解调器虽然属于通讯设备, 但由于它是用户实现联机必备的设施, 故而将其作为终端设备的一部分, 其作用是将计算机接收的指令处理好的机读信号, 转变为通讯设备和通讯线路可以接收和传送的电信号(这一过程称为调制过程), 然后送往主机系统。反之也可以将从主机返回的电信号转变成计算机可识别的机读信号(这一过程称为解调过程), 最后在屏幕上显示出来。

二、通讯线路及设备

通讯是实现联机检索, 特别是远程联机检索的关键。如果说终端设备相当于人的手, 主机相当于人的大脑, 则通讯就相当于为沟通大脑和手起传递信息作用的神经。

一般说来, 实现主机和终端之间的信息传递, 不但需要通讯线路, 还必须经过一个主要的分组交换数据的收集和发送装置, 这种设备通常设立在国家或当地电话、电讯部门。

在国际联机检索中, 主机和终端设备之间的信息传递除应具备

上述通讯条件外,还必须借助于国际通讯卫星。这是由于在国际联机时主机位于远隔重洋的外国领土,为了把情报传送给世界各地,主机必须与一个分组交换装置相连进入一个分组交换网,也称公用数据网。而终端设备与主要分组交换装置相连后,自然也要与主机所连接的分组交换网相连。沟通两个网络的媒介就是国际通讯卫星。

现以中国科学技术信息研究所(以下简称中信所)终端和西德STN主机系统的信息传递为例进行说明,中信所的终端设备是与我国的分组交换网 CNPACK(CHINA PACK)相连的。而STN系统主机则与世界著名的TELENET、TYMNET、COMPUSERVE网相连,其信息传递路径如图 1.3 所示。

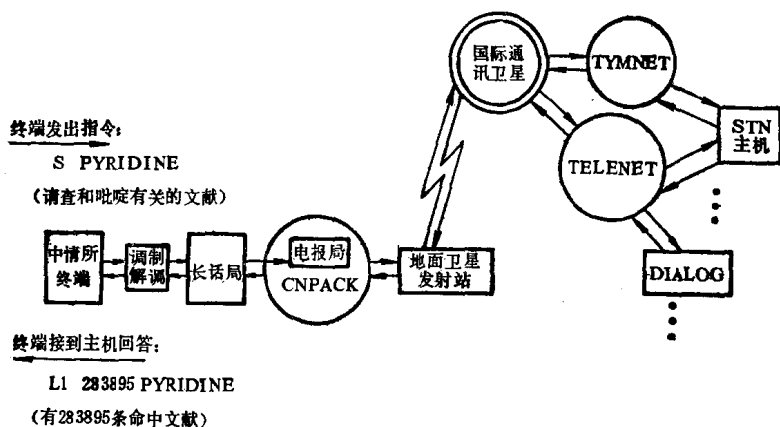


图 1.3. 国际联机检索信息传递示意图

在中信所终端只要输入 S PYRIDINE 指令,并按下回车键发出指令,信息便沿图 1.3 中上方箭头的方向传送到STN主机中。主机接收指令后回答的信息则以相反方向(如图 1.3 中下方箭头方向)送回中信所终端,屏幕上立即显示出用 PYRIDINE 标引的物质名称的记录命中 283895条。L1 是系统规定的文献集合编号,称为组号。

三、主机系统

通常我们称那些拥有大量数据库和各类情报数据的大型计算机系统, 及能进行情报管理(包括情报的收集、加工、生产及服务)的机构为主机系统。而我们经常提及的与主机系统联机, 是指与主机系统中的主要大型计算机(主机)联机, 文中为叙述方便也称与主机系统联机, 读者可按上下文内容判定。

主机系统是实现联机检索的重要条件之一, 获取大量情报的源泉。

主机系统数据库内所有情报数据的收集、组织和生产都是由像美国化学文摘社、英国皇家化学学会等出版商、数据库供应商完成的。他们将各种不同类型的图书、报纸、期刊、技术报告、会议论文、学位论文、专著、专利、标准、报表、目录、手册等刊登的信息, 收集、加工、编制成各种计算机可读形式磁带, 其中包括:

- 文献题录和文摘;
- 新闻公告;
- 统计表;
- 手册、指南、名录、商业财政数据;
- 日趋增加的全文文献。

然后由这些数据库生产商与主机系统签定合同, 将上述情报、数据以机读数据库形式装入系统的主机。当然主机系统也有自己生产的数据库。世界各地的用户可以通过通讯线路和网络与主机系统相连, 利用自己的终端或微机终端检索几十个或上百个数据库中存储的各类情报。

随着大型计算机、软件、通讯技术的飞速发展, 以及数据库生产数量的日益增长, 经营这种情报服务的主机系统也越来越多。据1990年的统计, 全世界大小主机系统约有655个, 而数据库多达4465个。世界最大的国际联机检索系统, 也是目前我们比较熟悉的美国DIALOG系统, 可检索的数据库有400多个。其次是德、

美、日联网合作的STN系统拥有100多个数据库(它是本书重点介绍的国际联机系统)。此外还有美国ORBIT系统、BRS系统,以及欧洲航天组织的ESA系统等。

四、数据库

从定义上来说,数据库是指存储在计算机中的机器可读、计算机可检索的情报集合体。而从人们可见的实体来看,数据库是由许许多多情报单元组成的,每一个情报单元通常称之为记录。数据库就是许许多多记录的集合,少则数千条,多则可达近千万条记录。

数据库生产商不同,其提供的数据库也不同。而各种数据库所提供的记录形式也可能不同。即便同一数据库生产商生产的数据库磁带,装入不同的主机系统也会出现不完全相同之处。用户关心的是数据库的内容,命中文献记录是否切合需要,以及查找原始文献。下面我们以联机检索STN系统的“美国化学文摘”数据库为例,介绍联机检出的记录形式。数据库记录通常有四种形式:(1)书目题录;(2)数值报表;(3)名录手册的事实型短文;(4)全文文献。现以书目题录形式的记录为例(图1.4)说明如下:

从以上STN系统CA数据库的记录实例来看,每一个记录内容包括若干个小的情报单元。例如TI代码右侧为文献题目;AU代码右侧为作者姓名等。我们把数据库记录中这种最小的情报单元称为字段,左侧的缩写字母为该字段的代码。每个记录就是由许多不同的字段组成的,它从不同的角度完整地反映了一篇文献的信息集合。例如,文献的题目字段(TI)表示文章的题目;作者字段(AU)表示文章作者姓名;CS字段是文章作者所在单位;LO字段是作者单位所在地;DT字段是文献类型;CO字段为刊登文章的期刊代码,IS表示期刊的国际标准期刊号;PY是文章发表的年代;LA是文章的语种;AB是文章的文摘;KW是关键词;IT是规范词等。

同一数据库的记录类型大体相似。例如上述记录是书目型文献记录,包括了题目、作者、出处、文摘、标引词等字段。但是,数

L2 ANSWER 1 OF 1 COPYRIGHT 1992 ACS
 AN CA112(11):98932J
 TI Stereochemistry of 3S-substituted alkoxyquinuclidine by
 two-dimensional and NOE difference NMR techniques
 AU Miao, Z. C.; Gao, J. H.; Feng, R.; Zhang, Q. K.
 CS Acad. Mil. Med. Sci.
 LO Beijing 100850, Peop. Rep. China
 SO Yaoxue Xuebao, 24(3), 194-9
 SC 31-4 (Alkaloids)
 DT J
 CO YHHPAL
 IS 0513-4870
 PY 1989
 LA Ch
 GI Diagram(s) available in offline prints and/or printed CA issue.
 AB The relation between the specific configuration of two diastereomers
 of the title compd. I and their biol. activities were studied by NMR
 spectroscopy.
 KW NMR two dimensional stereochem alkoxyquinuclidine; quinuclidine
 alkoxy stereochem NMR NOE
 IT Configuration
 Nuclear magnetic resonance
 (of alkoxyquinuclidine)
 IT Configuration
 (of alkoxyquinuclidine, NMR study of)
 IT 114121-70-9 114121-71-0
 (stereochem. of, NMR study of)

图 1.4

数据库不同, 其所提供的记录类型就不一定相同。目前若按所提供的记录类型来分, 数据库类型主要有 4 种: 文献书目型、数值型、事实型(或名录字典型)、全文型。本书介绍的 STN 系统“化学物质登记”数据库(REGISTRY) 是以名录字典型为主, 并加入了部分书目型记录, 其主要特点是大部分记录提供了图形, 它是一个集 4 种类型为一体、数据十分齐全、检索功能极强的数据库。

五、联机检索过程

本节所述联机检索过程是指选定检索主题词、编制好策略以后上机检索的基本过程, 包括上机检索的基本操作步骤。

1. 联机检索课题过程

用户带着准备检索的课题坐在终端机前, 利用键盘输入指令及

主题词向计算机提问，计算机检索系统可立即回答用户命中的文献量。这一问一答，一去一回的过程通常是在几秒钟之内完成的。

通过与计算机的问答过程获得检索结果，此结果就是命中文献记录的集合。输入显示指令，系统可显示 1 篇或 2 篇检出记录，从

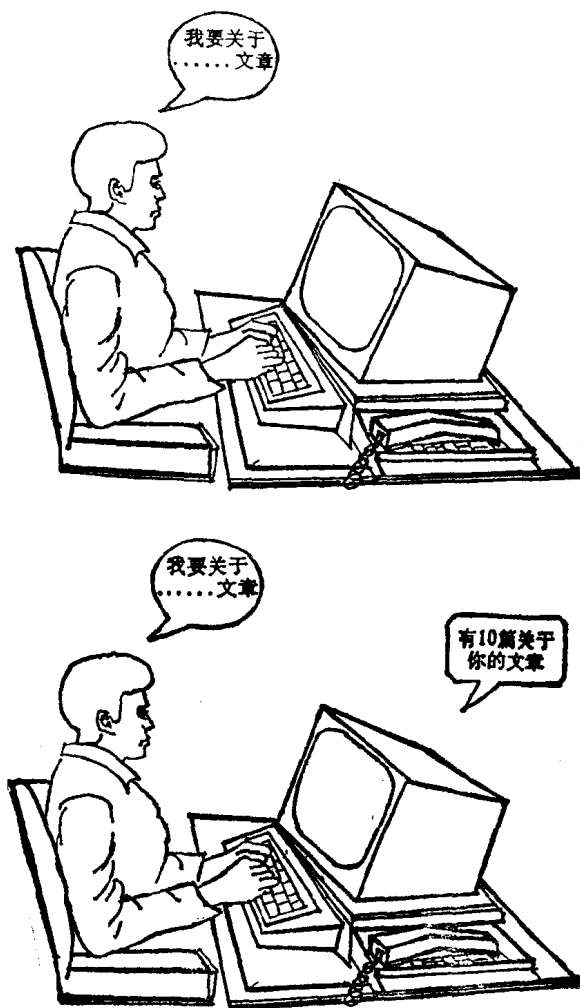


图 1.5

所显示的记录题目、文摘或标引词便可判断本次检索使用的方法是否正确。如果记录内容与用户要求不符,说明检索策略上存在一些问题。必须重选检索词或增加新的检索词,或重新组合概念。修改后的策略再用指令输入计算机,再次进行对话,直到检出结果令人满意为止。

如果用户对于检索结果表示满意,则可要求系统联机显示,并通过与终端相连的打印机随机打印出来,或者通过指令要求在主机系统的打印机上脱机打印出来,邮寄给用户。总的来说脱机打印的费用低于联机显示(打印),特别是打印数量比较大的情况下这种差别就更为明显。

2. 联机检索基本操作步骤

课题的检索过程通过如下五个联机基本操作步骤来实现:

第一步:接机过程;

第二步:进入数据库;

第三步:联机检索课题过程;

第四步:联机显示或脱机打印检索结果;

第五步:关机(或转入其它文档)。

下面分别介绍各个步骤的联机操作过程。(注意:在以下实例中,下划线部分均为用户输入部分)

第一步:接机过程

接机过程是在所有联机检索特别是国际联机检索中必须的一个步骤。接机是用户向终端输入密码的过程。就像打电话必须先拨电话号码,号码拨通了才能讲话一样。在联机检索时必须先向终端输入密码,密码输入后就连通了。主机系统显示欢迎信息后,用户就可以开始做联机检索课题。

国际联机接机过程中通常有4个密码是必须输入的:①主机所连接的网络地址码;②主机系统的标识码(常称主机标识码);③用户号(多指用户识别帐号);④用户保密密码(常称用户密码)。在这4个密码中有些系统不设主机标识码,而有些系统则用主机接口码来

代替。用户密码是用户在终端上进行联机检索的一把钥匙，有了它，用户才能进入主机。系统凭借用户密码和用户号计时计费，每月向用户邮寄帐单和收取检索费。如果用户密码失窃，则其它用户就可利用此密码检索课题及使用帐号。因此，用户密码一定要保密，一旦失窃可要求系统更改密码。图 1.6 是德、美、日 STN 系统的接机过程实例(注意：下划线部分是用户输入的各种密码)。

```

主机网络码--> 026245724720001
               CQM
主机接口码--> X
用户号--> LOGINID:rrrkalaad
用户密码--> PASSWORD:XXXXXXXXX

终端类型--> TERMINAL (ENTER 1,2,3, OR ?):3.

***** Welcome to STN International *****
               :
***** STN Karlsruhe *****
FILE 'HOME' ENTERED AT 06:51:17 ON 28 FEB 92

```

<—网络接通信息
 <—主机接通信息
 <—进入常驻文档

图 1.6

接机过程的详细说明见第三章第三节。系统接通后立即显示欢迎信息及系统新闻，然后转入价格较低的常驻文档(STN系统为HOME文档)，最后出现系统提示符。各个系统的提示符可能不相同，但都是提示用户输入指令的标志。必须记住，只有出现系统提示符后才可以输入指令。STN系统的提示符是箭头→。

第二步：进入数据库过程

这是选择文档，进入数据库的过程，需要用指令来完成。例如，STN系统如图 1.7 所示。

REG 是“化学物质登记数据库”(REGISTR) 文档的代码，FILE 是STN系统的选用数据库指令。

数据库的选择是联机检索首要和关键的步骤之一。如果说主机系统是一个大型的电子图书馆，则数据库就是不同专业书架上的各类图书。例如，我们要查找某一化学物质的情报，决不可能进入