第8卷 第4期 西北地震学报 Vol.8,No.4
1986年12月 NORTHWESTERN SEISMOLOGICAL JOURNAL Dec.,1986

地震学情报数据库研究

高树心
（国家地震局兰州地震研究所）

摘 要

将国内出版的地震学期刊论文，建立一个联机检索数据库。使用VAX11/750小型计算机和VAX DTR、CDD软件，操作系统是VAX/VMS。地震期刊论文数据库分成二个数据库，题录数据库和论文文摘数据库。它们是文件结构型数据库。一条题录记录占339字节，一条论文文摘记录占1290字节。自行编制设计了人机对话式联机检索程序。用户使用感到十分简便。论文追溯时间计划为十年或更长一些。

概 论

计算机情报检索系统，就是利用电子计算机，根据某种目的，在一定的时间内，从经过整理並已贮存在计算机内的情报中得到必要而充分的情报的系统。简单地讲，情报检索系统就是情报存贮和检索的技术。包括把文献加工为资料档的存贮技术和从资料档中找出所需文献的检索技术两个部分的内容。

电子计算机用于科学工程计算方面的系统，可以说是利用中央处理机运算功能的系统，即以“计算”为中心的系统，处理数据的资料档仅起暂存的作用；而情报检索系统却是以资料档为中心的系统，要把数据或各种记录作为资料档永久或半永久存贮在计算机中，供检索程序使用。一旦资料档的结构、形式和内容确定下来，检索程序的方法也大致确定下来。所以资料档在情报检索系统中占有很重要的地位。在各种情报检索系统中，文献情报检索最为重要，而且一般也比较复杂，因而研究得最多。

情报检索与数据库的关系最为密切。在现代情报检索系统中，数据库管理系统（DBMS-Data Base Management System）是它的最基本最核心的组成部分。数据库（Data Base）是计算技术的专用术语，是七十代国际上一个重要的研究课题。一个数据库可以定义为：一组存贮在一起，并具有尽可能小的冗余度的相互联系的数据，它以最优方式为一个或多个应用服务；数据被贮存得与具体的使用程序无关；使用一个公共的控制方案以增添新的数据，并修改和检索数据库内现有的数据。

计算机用于情报检索已有30年历史。它经历了脱机情报检索（1954～1964）；联机情

报检索（1965～1974）；和联机情报检索普及使用（1974年以后）三个发展时期。对用户一般提供三种服务方式：1.定题情报提供（SDI-Selective Dissemination of Information）.2.追溯检索（RS-Retrospective Search）.3.联机检索。前两种统称为脱机检索方式。地震学情报数据库的研究，就是以联机检索地震期刊论文数据库开始的。联机检索时，用户使用计算机终端直接与计算机对话，通过会话型检索程序进行“问答方式”提问並获得检索结果。检索地震学文献数据库时，用户调用检索程序，按提问要求键入检索项即可获得若干篇文献检索结果。也可先检索每篇文献的题录（控制号、年代、论文标题、出处、杂志代码、作者姓名、作者所属机构代码、分类号）。对中意的论文可通过控制号进一步要求它的详细情报—文摘。检索结果均可打印下来。如不合要求可另寻检索途径。

软件使用

地震学情报数据库的研究在VAX11/750小型计算机上进行，它的操作系统是VAX/VMS。85年8月本机安装了数据库管理系统(DBMS-Data Base Management System)；数据检索语言（DTR-Datatrieve）；公共数据字典（CDD-Common Data Dictionary）三个软件包。地震期刊论文数据库的设计使用了DTR、CDD二个软件包。

DTR语言是用于数据处理的一种有力工具。国外有人称为“咨询语言”，获得了广泛的应用。它是在COBOL语言的基础上发展起来的；但不要求数据库设计人员具备使用COBOL语言的知识和能力。它是一种第四代语言(fourth-generation language)。比COBOL，BASIC语言更加像英语（English-like）。它具有强的非过程特征。它吸收采用了数据库技术的一系列概念和方法，使数据库设计人员使用时感到十分方便，但亦不要求设计人员掌握数据库技术的知识和技巧。DTR提供了与其它语言相同的数据存贮能力，可以存贮和检索存在任何类型RMS（Record Management Services）数据文件中的数据，可以建立顺序文件和多个关键字的索引文件。

DTR可以存取三种不同类型的数据库.1.由DTR建立的文件结构型数据库；2.由关系数据库VAX Rdb（Relational data base）建立的数据库；3.由VAX DBMS建立的数据库。用DTR建库就是通过建造称之为域（domain）的库来存取数据。一个域的定义为一个数据的集合建立名称，並告诉DTR数据描述存贮在何处，数据存贮在何处。这样，一个域的定义就包含了域的名称，记录名称（数据描述）和数据文件名称。DTR也允许建立数据层次（如同COBOL中的组合项）和重复字段（如同COBOL中的OCCURS子句）。同时，DTR也为有经验的程序设计人员，准备了许多富有潜力的手段。通过编制“过程”的方式，在数据处理中高效率地使用计算机，以减少许多繁复劳动。

CDD软件实质是一个分层次的字典。字典是对CDD而言，目录是对VAX/VMS操作系统而言。当用户逻辑进入VAX/VMS操作系统时（DCL级$提示符），就处在某个缺省目录之下。同样，当运行进入DTR时（DTR提示符）就处在某个CDD的缺省字典之下。DTR就是使用CDD来贮存数据定义和过程。在CDD中各个数据定义与过程都是分开存贮的。在一个信息管理系统中，可靠的数据定义和数据本身一样重要。必须了解数据是如何表示的，当以不同的应用方式在系统中运行时它又是如何被使用的。共享的数据定义必须是无二义的，敏感的数据定义必须加以保护。CDD就是提供了这样一个中心存贮区域，並成为保

护数据定义的保密系统。数据文件並不包含在CDD中，而存放在VAX/VMS的目录之中，数据被贮存得与具体使用的程序无关。对数据文件的保护要在DCL级（$提示符）设置。用户可使用VAX/VMS建立新的目录，並在目录之间移动；同样，用户可以使用DTR建立新的字典，並在字典之间移动。

地震期刊论文数据库就是用DTR语言，建立的文件结构型数据库。但並不要求检索数据库的用户懂得DTR语言。用户只要调用数据库设计人员设计的问答式检索程序，简单地回答一些y（yes）、n（no），键入选用的检索项，就可获的检索结果。这好比並不要求收看电视的人懂电视机原理，只要简单地拨动设计人员为用户准备的几个旋钮，即可获得满意的图象一样。

地震学文献数据库

地震学文献数据库由三个数据库构成，一是带有文摘的文献数据库，域名是PAPERS；一是题录数据库，域名是SUBJECT；另一是文摘数据库，域名是ABSTRACT。计划收录范围为国内出版的全部地震学学术期刊论文，例如：地震学报、地震地质、西北地震学报、地震工程与工程振动、地震研究、地壳形变与地震等等，以及其它学科期刊中与地震预报研究有关的论文。追溯时间计划十年，还可以更长一些。

收集文献资料存贮到数据库中去是一项烦琐复杂的工作。文献是一种非数值情报，首先要对原情报进行加工、处理、形成可存贮到计算机内的二次情报、电子计算机检索到的二次情报一般可满足用户要求。若要看原文献，可根据打印下来的出处和刊名去借取。

利用每篇期刊论文的英文著录项目来建立地震学文献数据库。作为查找手段的检索项（年代、杂志代码、作者姓名、分类号）和用于输出的项目（控制号、标题、出处、文摘等）並列包含在文献数据库记录中。以下为一篇文献的记录形式：

CN : 850142
YEAR : 85
TITLE : High precision determination of the epicentral distance and azimnth angle
SOURCE : N 2 P133
CODEN : JSREAI
AUTHOR: Feng Rui
UNIT : 5
AUTHOR: Wang Bowen
UNIT : 5
CLASSIFY : 56.2578
ABS : Studies problems concerning the high precision determination of the epicentral distance and the azimuth angles. Suggested that within 3000 miles, Robbins formula should be used. Also furnishes corresponding FORTRAN program for the calculation.

控制号（CN）、年（YEAR）、标题（TITLE）、出处（SOURCE）、杂志代码（CODEN）、作者姓名（AUTHOR）、工作单位代码（UNIT）、分类号（CLASSIFY）和文摘（ABS）。其中控制号为主索引键，对每篇文献都不相同。作者姓名和工作单位代码是表字段AUTORS中的两个基本字段。根据国家规定作者最多收录两人。论文根据中科院图书分类法分类。工作单位使用了《地震文摘》中使用的两位符号单位代码，並做了大量的补充。一篇文献占1500字节存贮空间。

题录数据库记录形式如下：

CN　　：850142
YEAR　　：85
TITLE　　：High precision determination of the epicentral distance and azimuth angle
SOURCE　：N2 P133
CODEN　　：JSREAI
AUTHOR：Feng Rui
UNIT　：5
AUTHOR：Wang Bowen
UNIT　：5
CLASSIFY：56.2578

控制号（CN）是主索引键。文摘数据库记录形式为：

CN　　：850142
ABS　　：Studies problems conerning the high precision determination of the epicentral distance and the azimuth angles. Suggested that within 3000 miles, Robbins formula should be used. Also furnishes corresponding FORTRAN program for the calculation.

控制号（CN）即是主索引键也是检索项。文摘缩编到1000个英文字符以内。VAX11/750机每个输入行字符不得装过255个（包括空字符）。所以一篇文献的文摘用五个输入行组合构成。一篇文摘记录占1281字节。可以看出控制号是三个数据库相互关联的唯一纽带。

期刊代码即CODEN代码。CODEN是Code Number的简写，是计算机用的期刊代码。它是国际性的。目前国内使用尚不普遍，但随着计算机检索的普及，人们对它越来越熟悉。使用它可一定程度简化检索程序，也可促进国际交流。

CODEN规定用6个字符代表一个连续出版物的刊名。前五个字符为英文字母，最后一个字符为英文字或数字。前四个字母由刊名按一定的规则抽取而成，第五个字母是防止前四个字母重复而设置的。第六个字符为计算机校验字符，有一计算公式。地震学报、地震地质的CODEN代码分别是ASSID 7、DDIZD 4、与国际采用的代码相符。其余自行确定。

地震学文献数据库设计有两个字典表：期刊代码表（Coden-table；）和工作单位表（Unit-table；）。用户可使用这两个字典表，用一简单命令还原代码。例如

DTR〉print "JSREAI" via CODEN-table;

Journal of Seismological Research

DTR> print 5 via UNIT-table;

Geophysical Instituse, State Seismological Bureau

各种类型的情报检索系统有各自不同的检索语言和使用方法。我们的这个联机系统就是在VAX11/750机上使用DTR检索语言。虽然计算机提供了数据库软件,但各种应用程序—主要是检索程序－仍然要数据库设计人员自行编制。地震学文献数据库的总体结构请参看图1。

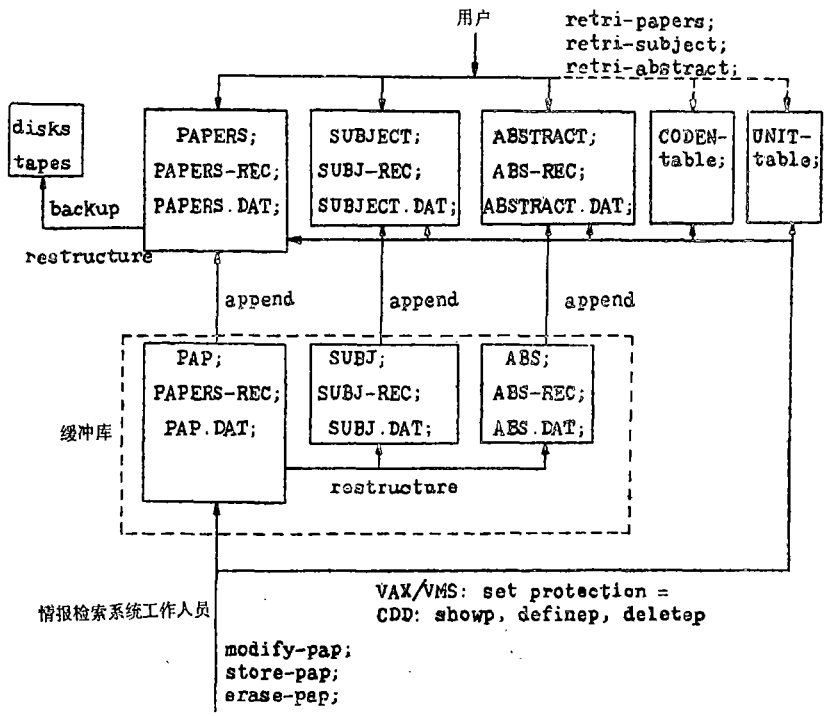

图1 地震学文献数据库总体结构图

Fig 1 Block diagram of seismological information database

应用程序和数据库设计如下:

- 文献数据库域的定义: PAPERS;
- 题录数据库域的定义: SUBJECT;
- 文摘数据库域的定义: ABSTRACT;
- 文献数据库记录定义: PAPERS-REC;
- 题录数据库记录定义: SUBJ-REC;
- 文摘数据库记录定义: ABC-REC;
- 文献数据库检索程序: Retri-papers;
- 题录数据库检索程序: Retri-subject;
- 文摘数据库检索程序: Retri-abstract;
- 三个缓冲数据库: PAP; SUBJ; ABS;
- 缓冲数据库存贮程序: STORE-PAP;
- 缓冲数据库修改程序: MODIFY-PAP;

• 缓冲数据库记录删除程序：ERASE-PAP；

• 二个字典表：CODEN-table；UNIT-table；

情报检索系统工作人员，使用缓冲数据库PAP；进行文献数据的存贮、修改、删除。检查无误后，再构（restructure）到另二个缓冲库SUBJ；ABS；再将三个缓冲库中的数据附加（append）到主数据库。每次附加工作结束后，将缓冲数据库对应的数据文件内的记录清除，等待输入又一批新文献数据。

用户使用的问答式文献检索程序框图示于图 2。调用检索程序实施检索进行情况如下：

```
DTR>: Retri-papers;
            Welcome to retrieve PAPERS data base.
        Answer following questions according to requirements!
Enter If use YEAR for retrieve (y or n): Y
Enter Please enter YEAR (2 digits): 85
    ⋮
Enter y for retrieving others, n to exit: n
Retri-papers program finished.Good-bye!
DTR>
```

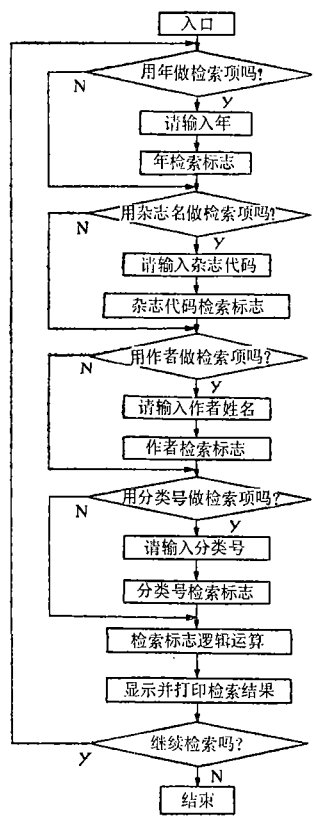

图 2　问答式检索程序框图

Fig. 2　Flow chart of a retrival program

地震学文献数据库设计中，在DCL级对数据文件设置了保护。利用CDD软件使用DTR检索语言对数据描述、过程程序和字典表设置了保密和保护。系统对数据文件提供缺省保

护：

```
$show protection
  SYSTEM=RWED, GROUP=RE,OWNER=RWED,WORLD=NO ACCE-
  SS
```

系统管理人员、文件主均有读（R）、写（W）、执行（E）、删除（D）四种特权，其它用户（WORLD）对文件没有任何存取权利。我们对用户检索的数据文件进行设置，例如：

```
$SET PROTECTION=(S: RWE, O: RWE,G:R,W:R) PAPERS.DAT;1
$DIR/PROT PAPERS. DAT;1
Directory SYS$SYSDEVICE: USER.GAO
PAPERS.DAT;1          (RWE, RWE, R, R)
```

对文献数据库文件PAPERS.DAT；1，所有用户都有读的权利，即检索的权利。並且都没有删除的权利。对用户使用的检索程序Retri-papers；设置的保密和保护如下：

```
DTR> SET DICTIONARY CDD$TOP.DTR$USERS.GAO
DTR> SHOWP RETRI-PAPERS; 1
1:    (300, 4),  USERNAME: "GAO"
      Grant-none, Deny-none, Banish-none
2:    (*, *),
      Grant-EPS, Deny-none, Banish-FG
```

结　语

文摘是文献最重要的情报，文摘的处理加工存贮到计算机中去，是一件极其烦琐的工作。根据计算机检索系统的要求和我们处理加工文摘的体会，对论文文摘的写作有一些想法需与广大地震论文作者商榷，拟另择文这里不再累述。仅希望我们建库工作能得广大论文作者的积极支持和密切配合。

国内目前大多数联机检索系统均为英文系统。我们在VAX11/750机上建的地震期刊论文数据库目前也是英文系统。检索试验表明：对于日常科研工作中，经常参考中文文献的同志，使用本检索系统，阅读本学科每篇文献的二次情报，不存在什么困难。当然随着VAX机中文系统软件的开发，将来地震学文献数据库应该是一个中文的联机检索系统。

（本文1986年1月收到）

参 考 文 献

〔1〕闻振远，电子计算机情报检索，人民邮电出版社，1981.
〔2〕H.S.希普斯，计算机情报检索导论，张承庆等译，知识出版社，1984.
〔3〕VAX Datatrieve User′ Guide
〔4〕VAX Datatrieve Reference Manual
〔5〕裴广生，小型情报检索专用数据库设计中的若干问题，计算机与图书馆，No.3，1983.
〔6〕顾成有，谈谈研究所图书馆自建文献库问题，计算机与图书馆，No.1，1980.
〔7〕涤非，连续出版物的CODEN代码，计算机与图书馆，No.3，1980.
〔8〕戎行等，SJTU科技情报检索系统，计算机与图书馆，No.2.1981.

A RESEARCH ON SEISMOLOGICAL INFORMATION DATABASE

Gao Shuxin

(*Seismological Institute of lanzhou, state Seismological Bureau, China*)

Abstract

Establishing an on-line retrievable journal paper database of seismology, these journals are published in China. Using VAX11/750 mini-computer, its operation system is VAX/VMS and VAX Datatrieve, VAX Common Data Dictionary software products. There are three file-structured databases, domain names are Papers, Subject Abstract, in this journal paper database of seismology. A record of Papers, Subject, Abstract is 1500, 225, 1281 bytes long, respectively.on-line retrieval programs are interactive mode, designed by myself. They are convenient to users. Retrospective time of this database is planning on going to ten years or more.