

北京气象中心实时气象资料库

应显勋 任满玲 葛 兰 李文华

(北京气象中心软件室)

提要 北京气象中心在 M-170 机上开发的实时气象资料库是一个功能较全、收集资料较多、实时性强、使用比较方便的专用气象资料库。本文对它的功能、设计思想、内部结构、程序结构及开发的技巧作了系统介绍。它可作为其他气象部门开发实时气象资料库的借鉴,也可使气象人员对利用数据库技术建成的气象资料有所了解。

北京气象中心自1980年建成北京气象通信枢纽(BQS系统)以来,实现了计算机通信和填图自动化,其后又开通了高速国际电路,沟通了同世界其他国家气象中心进行大批量气象数据的信息交换的渠道,每天收集到气象电报信息量达11兆字节。这些资料除向其他台站转发外,主要用于自动填图和客观分析,地面报和探空、测风报的标准层部分经规格化后分类存档,其他报告概以原始报文格式存档,其数据流程图见图1。

在 BQS 系统原设计方案中,没有资料检索系统,所以要在繁如浩海的原始资料中选取某些特定的信息只能分类打印输出,人工查找摘录,这种状况显然难以满足业务的发展以及科研的需求。

另外,近年来非常规资料的数量和质量都有提高,对非常规资料的利用及存档的要求也提出来了。为了更有效地管理组织,充分利用这些从全球气象通信系统(GTS)收到的大量资料,我们自行设计和建立了一个专用的实时气象资料库,资料库建成后,BQS系统数据流程如图2所示。

一、实时气象资料库的结构

本资料库是在 BQS 系统的 M-170 机上开发的一个面向气象业务检索、调用和存档的专用数据库系统,它是在虚拟操作系统 VOS2 的支持下,利用数据库技术,采用固

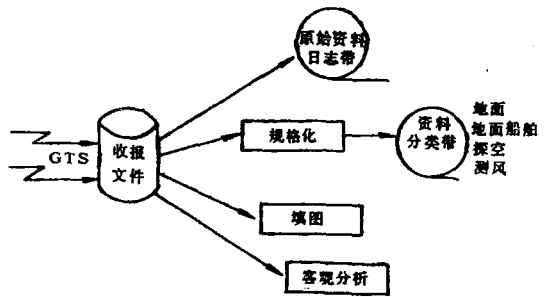


图1 原BQS系统的数据流程图

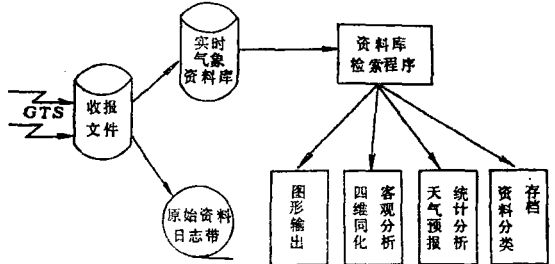


图2 目前BQS系统的数据流程图

定模式结构设计建立起来的。它共由四个子库组成。

1. 公报子库

公报子库是存放公报信息的资料库。它从通信线路上随机到达的各种点状分布的公报中提取报类(TT)、报头作为索引,以报头大小为序将公报存放到公报子库中。在处理过程中,一方面尽力保留资料的原始性,另一方面删除错误与重复的公报,使库内每份公报都能完整可用,随时满足用户的各种查询,较适合于气象电报的报类、报量

的统计、明码电报、台风警报的查询及确认电报到达与否等业务使用。

公报子库中收集的信息有：

报类	相应的 T T
① 国内旬月报	A B
② 地面分析报告	A S
③ 高空分析报告	A U
④ 其它分析报告	A X
⑤ 高空月平均报告	C E、C U
⑥ 洋面月平均和月总量报告	C O
⑦ 地面月平均和月总量报告	C H、C S
⑧ 国内各大城市预报报告	F S
⑨ 各种预报报告	FC、FJ、FR、 FT、FU、FX、 FZ
⑩ 格点值高度资料	G H
⑪ 格点值气压资料	G P
⑫ 格点值温度资料	G T
⑬ 格点值风场资料	G W
⑭ 日常航空报告	S A
⑮ 雷达观测天气报告	SB、SC、SD、
⑯ 地面观测报告	SI、SM、SN
⑰ 国内雨量报告	S L
⑱ 海洋深水温度观测报告	S O
⑲ 海洋温度盐度洋流观测报告	S O
⑳ 浮标站观测报告	S S
㉑ 海冰报告	S T
㉒ 湖冰报告	S V
㉓ 其他地面报告	S X
㉔ 气象卫星位置报告	T B
㉕ 气象卫星探测云资料报告	T C
㉖ 卫星探测晴空辐射报告	T R
㉗ 卫星探测高空压、温	T U

、湿报告

㉘ 卫星探测台风位置报告	T P
㉙ 卫星探测表温、风、云、辐射报告	T W、T S、T N
㉚ 非气象侦察机高空报告	U A、U T
㉛ 探空报告	U C、U E、U F、 U J、U K、U I、 U M、U N、U S、 U Y、U Z
㉜ 测风报告	U G、U H、U I、 U J、U P、U Q、 U Y
㉝ 气象侦察机高空报告	U R
㉞ 其它探空报	U X
㉟ 各种警报	W E、W H、W O W S、W T、W W

2. 报告子库

报告子库是建立在公报子库基础上的，存放以站为单位的报告信息的一个资料库。该子库对公报库中常用的并能进行数值处理的18类报进行加工处理。处理时，将每份公报里罗列的各站资料分门别类地提取出来，按一定的索引关系存入报告库。为了提高报告信息的准确度，对每份报告进行一系列的规格化检查，如时间订正、MiMiMjMj检测、字母数字变换、五码检查、指示码检查、重复订正处理等，排除了编报和通信传输中产生的错误。该库适合于资料交换、出版发行、历史资料存档和实时查询之用。

该子库所加工的报类如下：

- 第1类：地面报；
- 第2类：探空报A部和C部；
- 第3类：探空报B部和D部；
- 第4类：测风报A部和C部；
- 第5类：测风报B部和D部；
- 第6类：雷达观测天气报告；
- 第7类：卫星探测高空压、温、湿报告；
- 第8类：卫星探测晴空辐射报告；

第9类：卫星探测表温、风、云、辐射报告；

第10类：飞机探测高空报告；

第11类：卫星探测台风位置报告；

第12类：气象飞机探测飞行报告(UR中的一部分)；

第13类：人造探空观测报告(UX中的一种报告)；

第14类：浮标站资料报告；

第15类：国内旬报；

第16类：高空月平均报告；

第17类：暂缺

第18类：地面月平均报告；

第19类：国内各大城市预报报告。

3. 要素子库

要素子库是较报告子库更高一级的产品库。如果把报告子库中每一个具有气象含义的电码分离出来，根据世界气象组织发行的电码手册翻译转换成具有实际意义的气象要素值，那么这些要素值的集合就构成了要素子库。由于要素子库中的信息已完全变成数值型的，因此可为图形输出、统计分析、数值预报、气候评价和研究提供服务。为了提高资料质量，每个要素值在经过阈值检测和初步的相关性检查后，给出了一定的质量检测等级的标志，以供用户在使用时参考。要素子库的索引同报告子库。

4. 格点场子库

格点场子库是存放数值预报产品格点值的资料库。它把来自欧洲中心、华盛顿、东京的网格点报经格式化检查和译码后，拼成一个预报场或分析场存入场库，同时，提取日期、时次要素、地种位置层次、时效等关键字作成索引以使用户做各种选择。该库适合于各种物理量计算，有利于MOS预报和PP预报的开展及图形输出等业务使用。

二、实时气象资料库的特点

本资料库具有下列特点：

1. 实时性强

当BQS系统收到报文之后，本资料库系统能立即对其进行加工，存入相应资料库，供预报业务及其它用户使用。目前，由于M-170的资源(执行程序的分區)有限，故采用定时作业的方式，定时起动的间隔是可变的，可根据预报业务对气象资料的需要而定。

2. 信息量大

目前每天有11兆字节的资料加工入库，并供用户使用。详细加工的资料有18种，其格式不同，长度不一。各类报告中的数据属性也多，例如，一份报告中的一项数据与该报的报类、发报时间、地点、该数据的物理含义、质量状态等诸属性都有关联，这也说明资料库中关系复杂。

3. 集中管理减少了冗余度

本资料库根据各个用户的需要，从整体的角度来统一综合，并以数据库技术来描述、组织资料，从而实现了集中管理，消除了重复，减少了不必要的冗余。

4. 资料为用户所共享

本资料库有一套检索服务程序，用户可以通过这一程序，从系统控制台、终端、读卡机投入检索命令，或在程序中书写CALL语句，并填入相应的参数，就能获得各自所需的资料。在同一个时刻，可同时为几个不同的用户服务。

5. 资料的完整性

本资料库系统对存入库中的资料进行检测，在气象业务允许的范围内，对不合格式的数据进行更正，对不合气象含义的要素值不加更正，但给出其质量标志，从而保证入库的数据完整无缺、格式统一，提高了资料的可用度。

6. 资料的一致性

存入本库的资料，在复杂众多的联系中保持其一致性，不会出现矛盾。例如，昨天的资料不会出现在今天的索引之中；有了地面报的索引，必定存在地面报资料的实体；当指定某站的区站号或该站的经纬度后，所

索引到的必是该站的资料实体，等等。

7. 资料的独立性

本库的资料相对于用户是独立存在的，用户不必涉及资料库中的内部结构，只根据本身的需要来检索资料，资料库的用户界面（即用户程序和资料库的服务程序接口）保持不变，换句话说，不论资料库中内部结构发生怎样的变化，用户程序不必做任何修改。

8. 资料库的故障恢复

本资料库在运行中不论是正常结束或是出现了故障，在下次启动时都能从断点继续进行；对于发生故障的电报能自动跳过，从总体上保证资料库连续作成。

倘若由于某种原因，资料库遭到较大的破坏，则本系统能清除受破坏的部分，并利用 BQS 系统的收报日志带重新制作该部分，以保证资料的连续性。

三、实时气象资料库的服务

本资料库为用户提供四种查询服务方式：

1. 系统控制台命令方式

用户在系统控制台上启动资料库的查询程序，并投入有关的查询命令，查询到的信息将在宽行打印纸上输出。本方式可供机房操作人员对 BQS 系统收报情况进行监视或者其他人员索取报底之用。

2. 卡片输入批量方式

用户通过卡片机输入命令来检索资料库，结果将输出在宽行打印纸上或记入磁带。本方式只适用于公报库或报告库它便于在某段时间内固定地对资料库中某些内容检索。

3. 终端交互方式

用户在终端启动资料库终端检索服务程序后，对终端显示的信息进行应答，或投入查询命令，便可在终端屏幕上看到你所查询的数据。当查询到的数据太多时，你可中断显示，再次投入其他查询命令，进行其他数据的查询。本方式对四个子库都能检索，它便于预报人员随时、及时地调阅最新情报，

供预报分析之用。

4. 程序调用方式

用户在自己编制的程序里定义好必要的参数，然后利用 CALL 语句调用资料库的查询程序，其结果存放在用户指定的工作区内。除公报库外，其他三个子库都能利用本方式进行检索。它便于气象业务人员检索资料来直接进行加工计算，或记带归档。

本资料库中四个子库能提供查询的内容和范围如下：

公报库能为用户提供某段时间内所收集到的报类 (TT) 及报量，某种报类的报头或报文，指定报头中某几个字符的所有报头或报文。

用户可用查询命令，以经纬度、区站号或马士顿方格号三种不同的存取路径得到 18 种报类的离散点或连续区域中的报告信息。

对于要素库，用户可指定日期、时次、报类，以区站号或经纬度两种不同的路径，检索离散点或连续区域中某站报告中的某个段层某个要素或全部段层的要素。在终端交互方式下，还能提供在一定的区域中，满足某个要素的值域的站点及这个要素的具体数值。

对于场库，用户可指定日期、时次、要素、层次及预报时效来检索得到具体的格点场。在终端交互方式下，只能查询场库的索引，不能检索具体的格点值。

四、实时气象资料库的数据结构

数据是用来描述现实世界中事物特征的。一个事物的特征是多方面的，所以一般说来，描述一个事物的“数据”应由多个数据元所组成，数据结构是指数据或数据元之间的相互关系和约束，有时也指按某种关系结合起来的所有数据的全体。一个好的数据组织结构不仅有利于存贮空间的充分利用，提高设备的利用率，更重要的是还能使数据处理有较好的性能和实用性。

实用气象资料库要利用数据库技术来组

织数据，实际上就是要以数据库的观点来设计一个有效的数据结构，使得该结构能够：

- ① 准确地反映它所描述的对象，即事物及其相互间存在的某种自然联系；
- ② 条理清晰，便于存取；
- ③ 结构的次序不因存、删数据或数据元而改变；
- ④ 尽量以少量的存贮空间存放较多的数据。

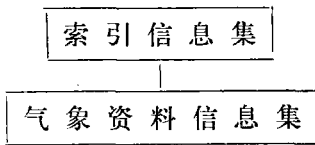
数据结构从其内容来看，可分为逻辑结构和物理结构。

逻辑结构是指组成结构的各数据之间的相互逻辑关系，是理论上的东西，反映在纸面上的。这种关系的表达称为数据模式，在逻辑结构设计中要说明数据元的格式、类型、取值范围和同其他数据的联系。

物理结构是指数据的逻辑在物理的存贮器上的映象，是具体的存贮结构。在物理结构设计中要指定各数据元及其联系在存贮空间中如何存放，这一般要考虑到机器字的大小、指令的动作、计算机能加工的数据格式、操作系统提供的存取方法等。

1. 实时气象资料库的逻辑结构

根据气象资料的特点及业务的需要，我们把本资料库的逻辑结构设计为如下形式：



索引信息集采用 1-n 的层次结构，其数据模式如图 3 所示。

我们把数据中的共同属性提取出来作为索引，它既反映了数据间的多种关联，又大大减少了数据冗余，并能满足用户的多种选择要求。

我们又把索引的层次设计得和存取路径相同，这样就能提高查询速度，另外，报头或经纬度这一级索引的数据量大，因此这级索引又采用了多级多分支树状结构，如图 4 所示。每个索引项节点下的分支数固定为 m ，

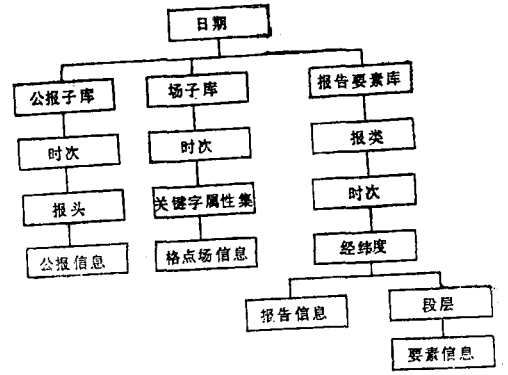


图 3 实时气象资料库数据模式示意图

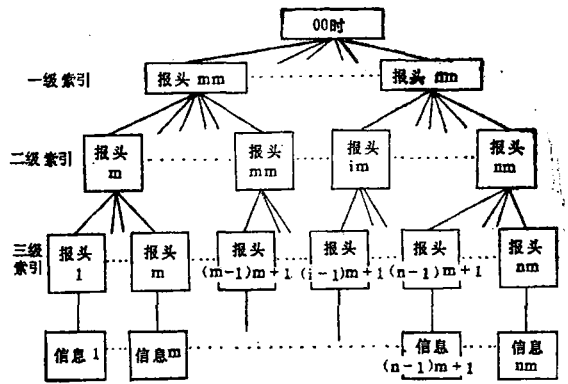


图 4 报头索引结构示意图

上一级索引节点上的报头或经纬度等于下一级索引分支中最后一个节点上的报头或经纬度，同一级分支中的报头或经纬度是从小到大排序的。

当然，有的用户需要以区站号来查询，但这种逻辑结构中没有区站号的索引，这是因为：

- ① 区站号只是经纬度的一个子集；
- ② 没有区站号的报类多于有区站号的报类；
- ③ M-170 的磁盘空间不很宽裕；
- ④ 建立区站号索引需增加资料库的生成时间；
- ⑤ 用户使用区站号检索少于使用经纬度检索。

但为之满足用户以区站号检索的要求，我们利用区站号数据字典，先从该字典中查

到经纬度，再从经纬度索引中得到具体资料。

2. 实时气象资料库的物理结构

为了使多个用户同时使用资料库，减少其等待时间，我们把上述的逻辑结构落实到几个文件，除了存放索引信息和资料信息外，还包含数据字典和文件管理信息，其结构如图 5。其中各文件存放的信息是：

HIF：日期总索引，公报库的时次、报头索引、HIF、IMF、GRF 的管理信息；

IMF：公报信息；

GRF：场库的关键字索引信息及场信息；

LIF：报告、要素库的报类、时次、经

纬度索引信息；

RMF：报告信息；

EMF：段层、要素信息及质量标志；

MCF：LIF、RMF、EMF 的管理信息，18种报类的电码数据描述字典；

SIF：区站号数据字典、马士顿方格数据字典。

这些文件都是直接存贮的固定长文件，文件之间、索引和信息之间的联系靠指针。同一级索引存放在同一个记录中，一个记录最多存放 m 个索引入口。对于不定长的资料由软件加工成固定长的后再写入文件。采用这样的形式能节省空间，便于存取，减少输入输出操作。

五、实时气象资料库的软件结构

本资料库系统由下列程序组成：

1. 初期状态设置程序

该程序的功能是分配磁盘空间，对文件进行初始处理，设置资料库运行时所需的各种初值，并作成数据字典。

2. 资料库生成程序

该程序的功能是从 BQS 系统的收报文件中读出资料，进行重报检查、规格化检查、译码和初步质量检查等预处理后，作成各子库的原始信息，然后按照资料库的数据结构建立各种索引，作成资料库。

该程序还负责资料库的资源管理、文件记录的动态分配。当资料库生成过程中发生异常时，能自动跳过故障信息，继续处理。

本资料库的数据流程见图 6。

3. 资料库清除程序

本程序的功能是自动删除资料库中最早一天的信息，回收文件空间，供新的信息存放之用；同时，生成资料库最新一天的日期索引信息。另外，还能清除指定日的信息和产生一个指定日期索引信息，以备恢复之用。

4. 资料库的恢复

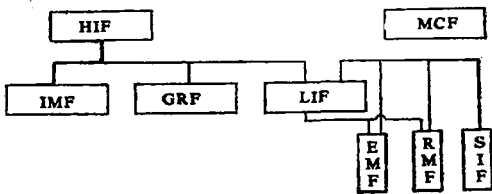


图 5 实时气象资料库的物理结构示意图

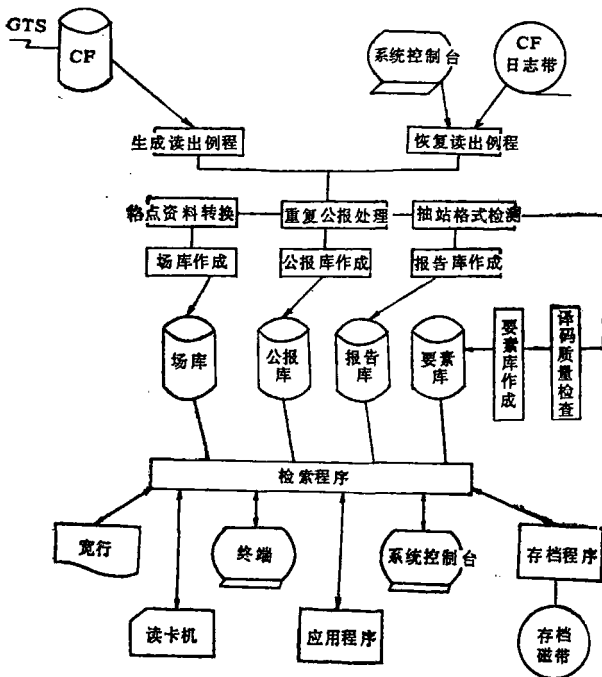


图 6 M-170 实时气象资料库总体流程图

本程序的功能是当资料库发生故障或遭到破坏时,利用 BQS 系统收报日志带作成指定日期的资料库或指定的子库。它还具备作成部分报类、部分子库的功能。因此通过本程序可将历史资料再现为资料库,以供需要历史资料的用户使用。

5. 资料库的检索程序

它由一套服务程序所组成,其功能是实现用户的各种检索,为用户提供服务。

6. 资料库的归档程序

本程序的功能是按照气候资料库的要求,将资料库中的信息分类记入磁带,为存档打下基础。

六、实时气象资料库的实现技巧

由于实时气象资料库的实时性强、信息量大、资料预处理复杂、磁盘空间有限等特点,因此在实现中要尽量考虑快速、节省空间、便于维护,为此我们采用下列技术方法来实现。

1. 动态生成索引

在资料库的索引中,日期、时次、报类等索引都是固定的因素,故我们采用了静态定义;但报头和经纬度不但量大,而且是随机到达,对于不同报类、不同日期、不同时间都是不相同的,故采用了动态生成的方式,以下通过报头索引来作一说明。

当一个新的报头来到时,先查看一级索引记录中的入口。当查到某个入口中的报头比它大时,就根据该入口中的二级索引记录的指针,把二级索引记录读出来,然后查看二级索引记录中的入口,若它大于所有入口中的报头,则把它作成该索引的最后一个入口,同时更新一级索引的那个索引中的报头;若它不是大于所有入口中的报头时,则把它作成一个中间入口插入到二级索引记录中,比它大的索引入口都往后移。

当一个二级索引记录中已达到 m 个入口时,则把该记录一分为二,并在一级索引中插入一个新的索引入口,当一级索引记录

中的入口也达到 m 个入口时,也可一分为二,并再产生新一级的索引,或者通过扩大 m 来解决。我们认为采用后一种方法好,目前在报头索引中 $m=70$,记录长为 1720 字节。

采用这样的方案,保证了上一级索引入口中,只对应一个下一级索引记录,从而减少了输入输出,便于快速插入及更新,于是加快了生成速度。

采用这样的方案,也便于查询。查询一次只需进行和索引级数相同次数的读入就能查到所需的信息,况且读入记录的入口是以大小排序的,故更有利于成批资料的检索。

如果采用某种固定方式的二级索引,就很难保证一级索引入口中,只对应一个二级索引的记录,于是增加了输入输出。我们曾经以报头的前三个字符作为一级索引,采用这样的结构比目前的结构在生成时,要多三分之二的输入输出。

2. 动态分配文件记录

BQS 系统每天收到的报量和每种报类的信息量都是不固定的,如果把磁盘空间以日期、时次、报类进行固定分配,则势必造成磁盘空间浪费。因而我们设计了一个与文件记录一一对应的位表。当某个子库需要空间时,就由动态分配例程从位表找到为 0 的位置成 1,并把相应的记录号记入记录号管理表,该记录就分配给那个子库了。这样,使文件空间为不同时次不同报类的资料所共享,起到节省磁盘空间的作用。

另外,在清除时,只要读出需要清除日期的记录号管理表,进行清除,并将其表中的相应位置置成 0 即可。不必把具体记录读出、清 0、写回,这大大节省了资料库的清除时间。

3. 采用数据字典技术

数据字典是用数据的语言来描述加工数据信息的一个特殊的数据库。它记录了加工数据元的名字、意义、来源、数据长度、变化范围、存取方法、加工方式、处理原则以

及和其它数据的关系。它是数据库管理中的一个重要工具。

由于气象资料格式比较复杂,变化频繁,故利用数据字典技术对气象资料进行分析、描述,有利于资料库的管理、软件设计及功能的扩充和故障的排除,同时也加快了处理和查询资料的速度。

4. 模块化程序结构

本资料库中的程序都采用了自上而下分层模块化程序结构,调用是单向的。一个模块完成一个功能,模块之间以接口联系。因此,本资料库的程序结构较清晰,易读易懂,也较容易编制,从而提高了可靠性,并便于维护。一旦发生故障,只要查阅有关的模块就能排除。当要扩充某项功能时,只要修改或增加有关的模块及接口即可。

本资料库共包含 144 个模块,约 30,000 条汇编指令。

七、实时气象资料库达到的技术指标

1. 信息量和空间

日处理的信息量为 10.5 兆字节, 20000 份公报。

公报子库约有 15000 份公报, 占空间 25 兆。

报告子库约有 50000 份报告, 占空间 10 兆。

要素子库约有 50000 份报告, 占空间 10 兆。

场库约有 98 个场, 占空间 0.5 兆。

实时气象资料库存放 5 天日报、5 旬旬报、5 月月报, 加上字典和控制信息, 约占磁盘空间 250 兆。

2. 实时性

生成一天资料库需 3 小时 10 分钟。

删除一天资料需 17 秒。

保存一天要素库资料(前 14 类报)需 21 分钟。

保存一天报告库资料(前 5 类报)需 21 分

钟。

在系统控制台上检索:

公报库 00 时次 约 4500 份公报需 5 分钟

报告库 使用区站号, 一份报告需 0.5 秒, 十份报告 2.5 秒, 6000 份报告约需 10 分钟; 使用经纬度, 检索前 14 类、00 时次的所有报告约 12000 份需 12 分钟。

用户程序检索资料库:

使用经纬度检索前 14 类 50000 份报告的所有要素需 20—22 分钟。

使用区站号检索 5 类 30,000 份报告的所有要素需 60 分钟。

检索场库一个场只需 0.07 秒。

终端检索资料库响应时间平均为 1—2 秒。

以上时间都是作业时间, 从中可看出, 用区站号来检索资料库是比较慢的, 这就是没有使用区站号来建立索引之故。它是通过区站号字典查到经纬度, 再以这个经纬度去检索资料库。如果检索程序能用较大的内存, 则也能加快速度。

八、实时气象资料库的效益

本资料库的建立将为气象资料的充分利用、远程终端的数据共享和实时预报业务的调用奠定了基础, 使业务预报、客观分析、图型输出、气象研究、气候评价、诊断分析等都能自动地查阅或使用资料, 免去了人工转换、读数、穿孔或终端录入等步骤, 给用户带来很大的方便。

目前北京气象中心的微机图型显示系统、人机对话系统(MCIDAS)、电视天气预报、MOS 预报、中期业务预报、某些非常规报的填图都使用了本库资料。另外, 科学院大气所和气象科学研究所的研究人员也来调用了资料库中的资料。我们还把本资料库移植到了 IBM 的 4381 机上为卫星气象中心的卫星资料反演提供服务。

本文曾得到许健民同志的指导, 并审阅了全部稿件, 我们在此表示感谢。