

基于 XML 和正则表达式的气象数据处理系统

田 兰 金石声 李 波 卜英竹 李 珏

(贵州省气象信息中心 贵阳 550002)

摘 要 随着气象综合观测网建设的不断推进,气象实时数据信息传输种类和数据量急剧增加。同时,气象实时数据信息的传输处理呈现某一特定时段需并行处理多类实时数据信息的特点。为实现气象数据信息高效、可靠、完整、及时的收集、存储和共享,采用 XML 标记语言对各类实时数据信息的报文内容进行标识和说明的方法,形成各数据业务类型定义,结合基于正则表达式的数据处理应用,使系统具备适应气象观测数据业务变化的能力。实现了对实时气象数据信息在省级中心的收集、处理、分发,改善了原有业务系统繁杂、功能单一、扩展性差、效率低等缺陷。

关键词 气象数据,XML,正则表达式,多线程

中图分类号 TP311 **文献标识码** A

Processing System of Meteorological Data Based on XML and Regular Expression

TIAN Lan JIN Shi-sheng LI Bo BO Ying-zhu LI Jue

(Guizhou Meteorological Information Center, Guiyang 550002, China)

Abstract With the continual progress of the construction of composite observing system for meteorology, the kind and amount of real-time meteorological data information transmission increase rapidly. Simultaneously, the transmission processing of real-time meteorological data information presents a characteristic that it needs the function of parallel processing of many kinds of real-time data information in a certain period. For gathering, storing, sharing the real-time data information efficiently, reliably, completely, rapidly, the technique that it adopts XML (Extensible Markup Language) to mark and explain the content of kinds of message information, and form the definition of data transmission operation, then it makes the system possess the ability that it can adapt the variable observing data transmission operation, with the data processing application based on regular expression. It realizes the gathering, processing and distributing the real-time meteorological data information in the provincial information center, and makes up the disadvantages of the original operation system that the system is complex, the function is tedious, its expansibility is not good, and the efficiency is low.

Keywords Meteorological data, XML, Regular expression, multithread

1 引言

近年来,气象综合观测网建设不断推进^[1],使得气象实时数据信息传输种类和数据量急剧增加,传输时效要求高,分钟级甚至秒级的数据越来越多,如多普勒雷达数据、闪电定位信息、区域自动站数据等。同时,预报预测业务系统、公共气象服务等系统,以及气象科学研究也要求获得更高效的气象数据服务。因此,大量实时气象数据,其高效、及时、完整、稳定、可靠、高质量地采集、传输、存储和共享是全国气象部门各级信息中心和相关业务部门的重要任务。

本文重点研究实时气象数据,它对短时临近预报、精细化预报、公共气象服务等实时业务非常重要。传统的实时气象数据处理系统由于不具备同时综合处理不同类型实时气象数据的能力,造成不同类型气象数据的处理需要专用的软硬件系统。随着各种新型气象探测手段不断得到应用,实时气象资料处理应用子系统越来越多,气象数据处理效率低、集约化

程度低的问题凸显。

基于 XML 和正则表达式的气象数据处理系统能够根据新增实时气象数据传输业务的要求,通过定义该类数据的业务模式,实现对该类气象数据的实时处理,有效改变了随着实时气象数据种类的增加,应用系统也不断增加的情况,同时利用多线程编程技术实现了实时气象数据的并行处理,较传统方式,实时处理效率得到了提高。

2 系统概述

基于 XML 和正则表达式的气象数据处理系统主要实现对气象观测台站按时效要求传输的实时观测数据的自动收集、处理、存储和分发。系统是在 vs. net2005 开发环境下使用 C# 编程语言,基于模块化编程开发的一套 c/s 应用软件。

2.1 总体功能结构

实时气象数据处理系统是气象信息业务平台中的一个关键节点。以贵州为例,各个气象观测站定时采集气象数据,通

过气象专用网络传输到省级气象数据收集服务器,运行在省级业务平台上的气象数据处理系统,将收集到实时数据分析处理后,提供给开展预报预测业务、公共气象服务业务等用户群体使用。

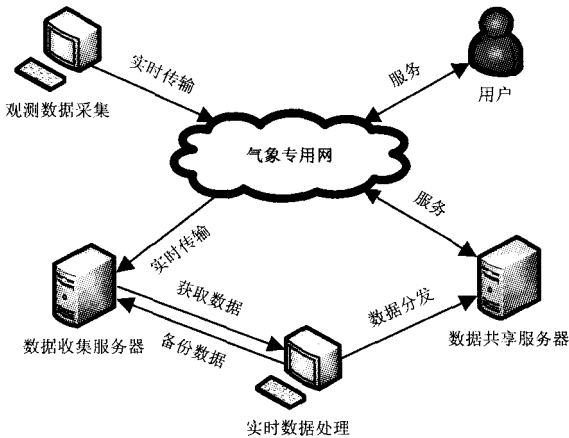


图1 总体结构

实时气象数据处理系统分为参数设置、系统监测、工具、外接程序等几个模块构成,具有外部程序链接、文件内容浏览、异常情况警示、定义新增实时气象数据业务、设置通信节点状态监测信息等功能^[2]。

“系统参数设置”菜单主要完成对各类实时观测数据的收集路径、分发目的地、数据格式等的定义。系统运行时先检查系统设置是否发生变化,发生变化就读取系统参数设置的配置文件信息,系统根据读取到的配置文件信息展示系统监测界面。

“系统监测”实现对各类实时观测数据的收集、处理情况的监测。系统完成初始化后缺省进入系统监测界面。

“工具”菜单主要提供“通信状态”、“人工处理”、“工作日志”、“命令提示符”等功能便于系统使用人员诊断和处理传输过程中的各种异常问题。

“外接程序”菜单主要实现外接程序可以调入对象模型以便实现应用程序自动化和扩展应用程序。

2.2 设计原则

系统开发中遵循的原则^[3]有:①灵活性和扩展性原则:随着气象综合观测网建设的不断推进,系统设计时必须能要能适应新型观测数据传输业务和原有业务的调整。因此,实时观测数据的传输目的地、数据类型、传输时间、数据文件规范及时效考核要求等要能进行动态定义。这样才能既满足现有实时气象观测数据传输业务的运行需要,又能适应今后业务的发展。②及时性原则:气象预报预测业务、公共气象服务业务等对实时观测数据的时效性要求越来越高,除了新增观测数据种类逐年增加外,原有实时数据收发的时间间隔在不断缩短、数量却呈现不断增加趋势。在系统运行稳定的基础上快速处理,才能适应越来越密集的数据处理要求。③稳定性原则:各类气象实时数据信息的传输业务是全天候不间断运行的,因此气象实时数据信息处理系统的稳定性是决定系统成败的关键因素。

2.3 处理策略

实时气象数据的处理策略将决定整个系统的性能。绝大

部分的实时气象数据均是按一定的时效要求定时传输的,例如雷达扫描数据每6分钟生成一次,国家级自动站观测数据每小时传输一次。目前仅气象预警信息是依据灾害性天气实况随机生成的。

处理这些气象数据时,要实现动态定义实时气象数据类型就必须解决将晦涩难懂的气象数据格式(气象系统内部的实时观测数据采用的是“五位摩尔斯电报码”进行的编码)转换为便于系统用户理解和修改的文档。系统采用了XML标记语言定义配置文件,守护进程依据配置文件设置各类气象数据的传输时效定时调用相应的线程进行实时处理,各线程之间是并行运行。处理过程中为提高效率,数据文件匹配和字符操作等采用了正则表达式。

通过这种处理策略的设计,既提高了系统的灵活性、扩展性,又节约了系统资源,提高了数据处理速率。

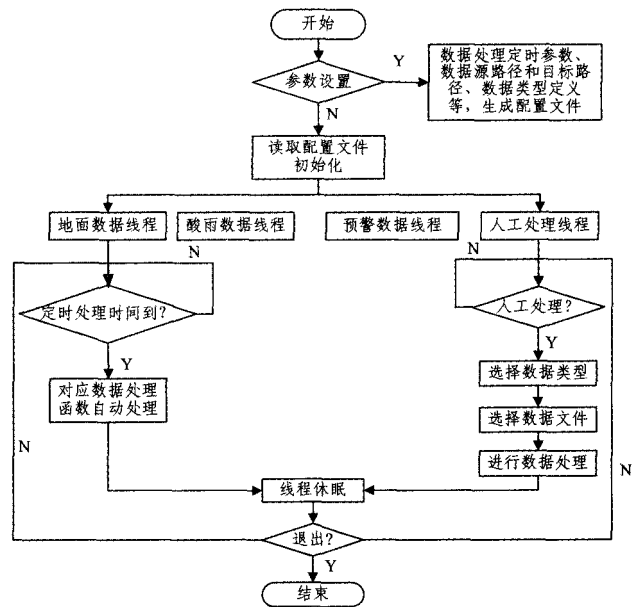


图2 处理策略流程

3 系统技术

气象实时数据信息的收集情况、传输节点通信状态监测和实时数据信息处理情况的动态展示是该系统的关键性的功能,也是整个开发过程中的难点和重点,包括了气象实时数据信息文件分类和类别描述、文件名规则分析、存储结构目录设计、文件格式分析及气象实时数据信息传输流程分析和传输方式分析。

3.1 采用XML定义配置文件

参数配置^[4]是系统初始化的依据,为了适应新增气象观测数据类型和现有气象观测数据业务流程调整的情况,系统必须实现能通过动态的配置方式来改变处理方式。该系统采用XML标记语言对气象观测站上传的各类实时气象数据信息的内容进行标识和说明,形成了对该类实时数据传输业务的定义。因此在系统开发的过程中,使用XmlDocument生成配置文件来保存系统参数和运行状态数据,通过C#直接操作XML数据。以酸雨观测资料的实时传输处理为例,通过定义该类数据信息的处理方式、处理频次、处理时间、源文件来源和文件名格式、及文件中报头报尾标识符等信息生成酸

雨观测数据传输业务模式。系统初始化时先读取 sys.xml 文件中的内容,确定系统监控显示界面,再读取 FILERCV_DIC.xml 文件中对各类实时报文资料传输业务的定义生成当天系统运行的节目表,从而实现对各类实时传输资料的收集、处理、传输和分发。

3.2 采用正则表达式进行文件匹配和字符操作

该系统运行过程中,对于实时气象数据的识别,处理方式等的判别多依赖于对该类实时数据信息文件名或内容的识别。为了提高效率和可靠性,正则表达式^[5]是一种非常有效和高效的处理方法,可以被看成是一个强大的通配符(通用匹配符号),但它提供了比这种通配符能力更强、控制规则更复杂、功能更完善的匹配机制。正则表达式(regular expression)是用来快速、高效地处理文本数据的工具。系统通过在 XML 配置文件逐一类每一类实时数据文件,采用正则表达式对文件命名、文件内容、格式等进行模式定义。对各类收集的实时数据信息文件名和文件内容、格式等与配置文件中预定义模式进行匹配,从而实现快速、准确的数据检索和提取。

例如,地面数据文件名的正则表达式为:

```
string Pattern=@"^SP((0[1-9])|((1|2)[0-9])|30|31)((0|1)[0-9])|(2[0-3])((0|1|2|3|4|5)[0-9])\.([0-9]{3})$";
System.Text.RegularExpressions.Regex reg = new System.Text.RegularExpressions.Regex(Pattern);
Bool match=reg.IsMatch("SP100600.606");
```

系统利用.NET 框架中处理正则表达式的关键类 RegEx 类,该 RegEx 类包含了一个名为 IsMatch 的静态方法,它返回一个布尔值,这个布尔值说明指定的输入串是否与一个给定的正则表达式匹配。这样会比从文献中截取字符逐一对比的方法在功能和效率上要高很多。该系统在实时数据文件内容判别和操作上都采用了这种技术。

3.3 自动处理采用多线程处理机制

由于该系统涉及的实时数据信息类型比较多,且呈现某一特定时段需同时处理多类实时数据信息的特点。为保证数据处理和传输的有效性和及时性,使用了多线程技术^[6]来解决。

系统为每一类实时数据信息建立一个线程,每隔一定的时间间隔对配置文件进行扫描,将到达收集、传输时间的记录提取出来,逐个搜索符合条件的实时数据文件,当存在条件相

符的文件时,则通过读取系统日志文件判断是否需要处理。比如,高空探测数据的处理过程:

```
private Thread thread_GKB;
private void GKB_Thread()
{
    while (thread_GKB.ThreadState == System.Threading.ThreadState.Running)
    {
        GKB_Process();
        Thread.Sleep(60000);
    }
}
private void GKB_Process()
{
    SendInfo SendInfoList=GKB.Send(s_host,s_userid,s_password,t_host,t_userid,t_password,s_copyto_network,s_copyto_network_userid,s_copyto_network_password,s_error_path,sys_station_code,sys_station_name);
    listSuccess(SendInfoList,GKB.Success_listView);
    pri(SendInfoList,GKB.Error_listView);
    listProcessLog(SendInfoList,ProcessLogInfo,ProcessLog);
}
```

同时,如果不采用多线程,那么在系统进行实时数据的自动处理过程中,系统窗体基本是“死”的状态,将无法进行人工操作。所以系统采用多线程操作也是必须的。

4 系统的业务实用性

基于 XML 和正则表达式的气象实时数据信息处理系统着眼中国局 2012 年 4 月投入业务运行的“新一代国内气象信息系统”,旨在优化业务流程,改善现有省级气象数据处理系统无法扩展,不能适应新增观测数据业务或现有数据业务发生改变的情况、补丁程序多、异常情况时有发生的情况。系统自投入业务运行以来高效稳定、操作简便、监测信息准确清晰、界面友好。业务应用表明,该系统易于学习和掌握,易于维护和扩展,具有很好的实用性和通用性。

采用新方法开发的气象实时数据处理系统相比传统应用系统具有明显优势,如表 1 所列。

表 1 新旧系统应用对比

需运行应用系统数	是否提供初级质量控制提示信息	是否能便捷实现新增观测数据业务	现行观测数据业务发生变更时是否需修改程序	是否具备可扩展性	是否提供各类数据的收集情况监视
新系统	1	是,提供针对文件名、编发时次、报头和报尾等初步质量控制信息	否,只需修改配置文件	是,具备外接程序功能	是
老系统	7	否,要开发新系统	是,需要修改源程序	否	仅有几类数据收集情况监视

结束语 基于 XML 和正则表达式的气象数据处理系统采用多线程的编程技术,实现了地面气象观测业务调整后各类实时数据信息的收集、处理和分发,为适应综合气象观测系统新增数据业务奠定了基础,具备很好的业务推广前景,具有以下特点^[7]:

①采用了 XML 标记语言来定义数据传输业务,增强了

数据处理过程中的准确性。

②系统能通过修改配置文件的方式实现新增数据业务,具备可扩展性。

③系统具有友善的监控界面,错误信息更加容易定位。

④系统具备对各类数据收集和处理情况的日志记录。

该系统在气象数据种类和数量不断增加的环境下,能够

对不同种类气象数据进行实时处理,并且实现了对各类数据动态配置的功能。

参 考 文 献

[1] 中国气象局监测网络司. 气象信息网络传输业务手册[M]. 北京:气象出版社,2006
 [2] 孙林华,惠志红,邵亮,等. 基于 SCO Unix 的省级气象信息编解码处理系统[J]. 气象科技, 2007, 35(3):445-447
 [3] 钟艳雯,罗小珠,夏正龙,等. 湖南省级气象信息传输系统研制与开发[C]//2011年中国气象学会气象通信与信息技术委员会暨

国家气象信息中心科技年会论文集. 北京:中国气象学会,2011
 [4] 王甫棣,姚燕,李湘. 基于 XML 的气象数据订阅系统设计[J]. 气象科技,2012,40(4):591-595
 [5] 管杰裕. 正则表达式在气象信息处理中的应用[J]. 广西气象, 2006,27(增刊1):107-111
 [6] 薛建军,周杰,杜景林,等. 基于多线程的实时与非确定时气象资料处理[J]. 信息技术,2011,11:19-23
 [7] 王丽政,任晓炜,李涛. 广西气象信息网络传输业务实时监控系统的设计和实现[J]. 气象研究与应用,2011,32(增刊2):274-277

(上接第 427 页)

AGENTID	DIRECTION	TALKDUR	ANSWERTIME	ISCONN
10000001	呼入	120	10	是
10000002	呼出	150	15	是
10000003	呼入	180	20	否
10000004	呼出	210	25	是
10000005	呼入	240	30	是
10000006	呼出	270	35	否
10000007	呼入	300	40	是
10000008	呼出	330	45	是
10000009	呼入	360	50	否
10000010	呼出	390	55	是

图 4 报表显示

(2) 呼叫数据报表

呼叫中心收集到的信息全部汇集到呼叫数据报表中,如果需要某些数据,或定制某些需求,直接点击图 5 所示的红色区域“定制指标”后就可进入到报表定制开发界面。

日期	呼入	呼出	呼入呼出	呼入呼出	呼入呼出	呼入呼出	呼入呼出	呼入呼出	呼入呼出
2012-12-21 00:00	2008	86	508	243	26	0	0	0	0
2012-12-21 01:00	1762	615	1147	1147	20	20	100	100	100
2012-12-21 02:00	1875	688	1187	1187	16	16	19	92.75	92.75
2012-12-21 03:00	1296	217	1513	1513	28	28	22	84.65	84.65
2012-12-21 04:00	1476	626	1100	1100	17	17	17	100	100
2012-12-21 05:00	1410	871	539	539	14	14	14	100	100
2012-12-21 06:00	1107	841	266	266	5	5	5	100	100
2012-12-21 07:00	690	841	150	150	14	14	14	100	100
2012-12-21 08:00	876	202	1078	1078	13	13	13	100	100
2012-12-21 09:00	1466	204	1262	1262	14	14	14	100	100
2012-12-21 10:00	1382	223	1159	1159	14	14	13	92.86	92.86
2012-12-21 11:00	1506	289	1217	1217	5	5	5	100	100
2012-12-21 12:00	2111	211	1900	1900	8	8	8	100	100
2012-12-21 13:00	1234	605	629	629	15	15	15	100	100
2012-12-21 14:00	1438	751	687	687	22	22	20	90.91	90.91
2012-12-21 15:00	1987	627	1360	1360	23	23	28	96.55	96.55
2012-12-21 16:00	1614	600	1014	1014	21	21	18	78.56	78.56
2012-12-21 17:00	2043	982	1061	1061	26	26	26	100	100

图 5 呼叫数据界面

双击进入定制指标进行报表定制开发,如图 6 所示。

报表名称: 危险外呼组坐席外呼报表
 指标名称:
 是否审核
 是否存储过程
 指标描述: (危险外呼组坐席外呼报表,日报表)
 报表名:
 筛选条件: ISF_CALLDUR_IAMTSTA
 临时报表名: ISF_CALLDUR_IAMTSTA
 是否定制指标
 定制输入参数:
 是否第三方指标
 第三方页面URL:
 可以进行的周期统计: 按天统计 按周统计 按月统计 按分钟统计

图 6 报表定制开发界面

若用户需要定制的报表功能较复杂,也可直接将已保存的存储过程进行调用,从而定制不同类型的报表。其存储过

程调用如下:

```

From(
Select(
Agtd,startdate,
Sum(talkdur) as sumtalkdur,——通话总时间
Sum(case when(direction='呼入'or direction='呼出') then 1 else 0
end) as sumcallout,
Sum(case when isconn='是'and (direction='呼入'or drection='
呼出')then 1 else 0 end) as sum
Round(convert(float,sum(talkdur))/sum(case when isconn='是'
and (direction='呼入'or direction(sum(talkdur)+sum(answer)) as
callout——总外呼时长
From
App_inf_agtcall_dtl_raw a
Where
startdate between @start_date and @end_date
group by agtid,startdate
) aa where #t_callout_shengming.agtid=aa.agtid and #t_callout_
shengming.startdate=aa.startdate
    
```

结束语 综上所述,SOA 技术应用在呼叫中心系统报表制作方面,实现了报表模块与原信息系统的松耦合,增强了系统灵活性。以后将利用业界新技术,不断完善平台系统,使得未来的呼叫中心报表制作可以人性化地通过界面上的简单配置来实现,从而逐步实现基于 SOA 服务架构能全面面向报表制作人员服务。

参 考 文 献

[1] 呼叫中心-电讯在线[OL]. <http://www.call-center.net.cn/>
 [2] 叶庆卫. 基于 HTML 的报表系统研制与开发[J]. 计算机应用研究,2000(12)
 [3] 徐雅斌. 基于 CTI 技术的呼叫中心的设计与实现[J]. 计算机工程,2007(5)
 [4] 许妮. 呼叫中心的核心技术及组成[D]. 武汉:武汉大学信息管理学院
 [5] 王磊. 基于 SOA 的报表服务模型的研究与设计[D]. 上海:上海交通大学,2009
 [6] 赵延超. 呼叫中心中可定制报表系统的设计与实现[D]. 西安:西安电子科技大学,2012