

基于表驱动的动态数据上报管理系统研究

李林¹ 郑海宁¹ 彭帆¹ 顾进锋¹ 陆书涵² 张大红³

(1. 中国农业大学信息与电气工程学院, 北京 100083; 2. 俄亥俄州立大学计算机与信息科学, 哥伦布市 43210;

3. 北京林业大学经济管理学院, 北京 100083)

摘要: 近年来,信息管理系统广泛应用于各个领域,由于大多研发者只着眼于设计开发自己所需求的应用系统,由此产生一系列问题:各领域内系统整体开发效率低下,系统的扩展性和通用性差,造成了资源的大量浪费。本文针对数据填报及统计类系统领域,设计并实现包含数据访问构件、数据校验构件、空缺数据填补构件和统计结果查询输出构件在内的可复用软件——基于表驱动的动态数据上报管理系统(DDTMS),以解决信息管理系统中数据填报及统计类系统的共性问题。将DDTMS应用到中国林业生态资源环境承载力:数据投放器和中国生态安全指数系统中,试验表明:该可复用软件具有良好的通用性,极大提高了系统开发效率,改善了系统的扩展性,表现出良好的实用性和稳定性。

关键词: 表驱动法; 可复用软件; 动态编译; 数据填补

中图分类号: S2 **文献标识码:** A **文章编号:** 1000-1298(2017)12-0190-12

Dynamic Data Transmission Management System Based on Table-driven Methods

LI Lin¹ ZHENG Haining¹ PENG Fan¹ GU Jinfeng¹ LU Shuhan² ZHANG Dahong³

(1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

2. Computer and Information Science, University of Ohio State, Columbus 43210, USA

3. School of Economics and Management, Beijing Forestry University, Beijing 100083, China)

Abstract: In recent years, information management system (IMS) has been widely applied to every different fields. A series of problems arose because most developers only focused on the design and implement of their own application systems. It caused resources to be wasted in various fields. Meanwhile, the system development efficiency was low, and the expansibility and versatility of system were poor. The object was to design and implement a reusable software system to meet the common demands of information collection and statistic system in IMS. The dynamic data transmission management system (DDTMS) was based on table-driven methods, and it contained data access component, data verification component, vacancy data filling component and statistical result query component. Each component performed specific functions. The dynamic data transmission management system was developed on Visual Studio 2013 platform. Lots of technologies were used to finish the reusable software system like factory pattern, dynamic compilation, jQuery and C#, and so on. In order to test the data transmission management system sufficiently, it was applied to two systems. One system was Chinese forestry ecological resources environmental bearing capacity: data dispenser, and another system was China ecological security index system. The experimental results showed that the reusable software system was practical and stable. It was of good versatility. And it had greatly improved the development efficiency and the expansibility of system.

Key words: table-driven methods; reusable software; dynamic compilation; data filling

收稿日期: 2017-04-14 修回日期: 2017-05-23

基金项目: 国家自然科学基金项目(31471762)

作者简介: 李林(1963—),女,教授,博士生导师,主要从事软件工程和软件自动化研究,E-mail: lilincan@126.com

通信作者: 张大红(1959—),男,教授,博士生导师,主要从事林业技术与经济研究,E-mail: zhangdahong591120@163.com

引言

随着计算机信息技术的飞速发展,信息管理系统广泛应用于各个领域^[1-6],相比于传统人工管理方式,使用计算机进行信息管理具有无法比拟的优势:数据处理效率高、数据查询快速准确以及信息便于长期保存等。信息管理系统种类较多,包含办公自动化系统、通信系统、交易处理系统、决策支持系统和数据填报及统计系统等,其中数据填报及统计系统在生态环境监测和农业生产信息管理中广泛应用^[5-9]。

数据填报及统计系统的一般业务流程为:填报人员填写并上报数据,在数据中心集中进行数据管理、统计和计算工作,统计与计算的结果供各级用户查询浏览。在传统的数​​据填报及统计系统实现过程中,存在许多实际问题需要解决:反复从零开始,开发各种不同应用目的的数据填报及统计系统,使得单个系统的开发速度慢、周期长,开发效率低下,同时造成数据填报及统计类系统领域中代码冗余量很大;将已完成的系统更换不同数据库类型运行,需要大量修改程序,系统本身无法灵活适应不同的数据库环境;当系统需求改变时,需要不断修改代码和测试程序,系统维护困难,不易扩充。

表驱动法是改善系统可维护性和提高系统可扩展性的有效手段之一,研究者多建立在具体需求结构化分析的基础上,对其研究以简化复杂任务,提高工作效率^[10-12]。软件复用是提高软件开发效率的通用思想,也是提高软件可维护性的高效手段。可复用软件的应用可极大降低相关领域内软件重复设计带来的资源浪费,提高同领域内软件的规范化和标准化^[13-17]。

本文针对传统的数据填报及统计系统实现过程中存在的问题,在表驱动思想下,进行相关数据库设计,并在此基础上设计可复用软件——动态数据上报管理系统,完成基于表驱动的动态数据上报管理系统,并将其应用到中国林业生态资源环境承载力:数据投放器和中国生态安全指数系统中,以测试该可复用软件的有效性和实用性。

1 数据填报及统计系统

在数据填报及统计系统中,数据的输入、运算和输出是其基本功能。在数据输入时,数据校验是重要的基础性功能;进行统计分析时,填补缺失数据是首先要解决的问题;统计分析后,统计结果查询输出是必不可少的部分。为此,设计基于表驱动的 DDTMS,可以解决信息管理系统中数据填报及统计

类系统的共性问题。

数据填报及统计系统中基于表驱动的 DDTMS 与应用间的关系如图 1 所示。数据层的指标结构表是数据库的核心,包含系统所需收集的指标体系,是指标和指标基础数据以及指标和指标填报数据间的桥梁。而统计项结构表是系统的辅助结构表,包含系统所需统计的内容项,是统计项和统计数据间的桥梁。两个结构表共同驱动上层的 DDTMS 完成其相应功能。DDTMS 位于中间层,受数据层 2 个结构表驱动服务于应用层中的各应用。基于表驱动的数据上报管理系统动态提供数据访问、数据校验、空缺数据填补以及统计结果查询输出功能,通用于数据填报及统计系统中。

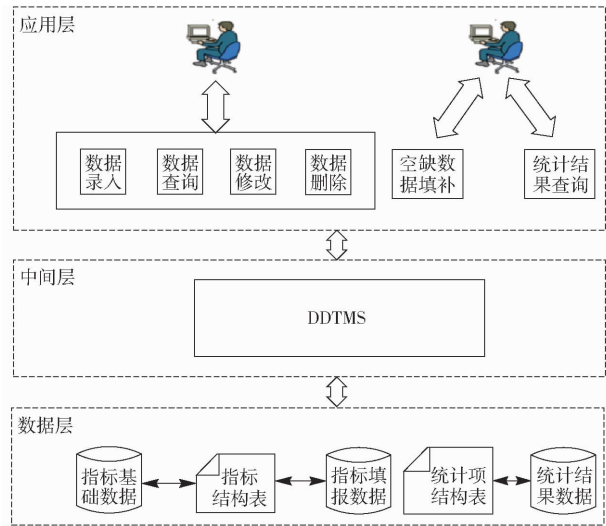


图 1 DDTMS 与应用间的关系示意图

Fig.1 Sketch of relationship between DDTMS and application

2 表驱动及数据库设计

在不同的数据填报及统计系统中,所需收集的指标和指标描述参数不同,并且需提供具体选项以供用户选择的描述参数在不同指标间存在差异。将指标及其相关数据校验和空缺数据填补通过具体的程序实现,同时将指标、指标描述参数和具体选项内容以固定的格式和内容在数据填报及查询界面中显示,开发效率低,同时系统难以灵活适应各种条件的变化,在不同的数据填报及统计系统间更缺乏普适性。

针对上述问题,基于表驱动法^[10-12]设计可扩展数据库^[18-19],表间关系图如图 2 所示。其中 T_ZBB 是指标结构表(以下简称结构表)。结构表是建立在具体需求结构化分析的基础上,形成的指标与不同功能中具体信息的对应关系表,是系统的驱动表,也是整个数据库的核心表。T_ZB0_LY、T_ZB0_DW 和 T_ZB0_KQ 这 3 个表是基础数据表,分别存放供

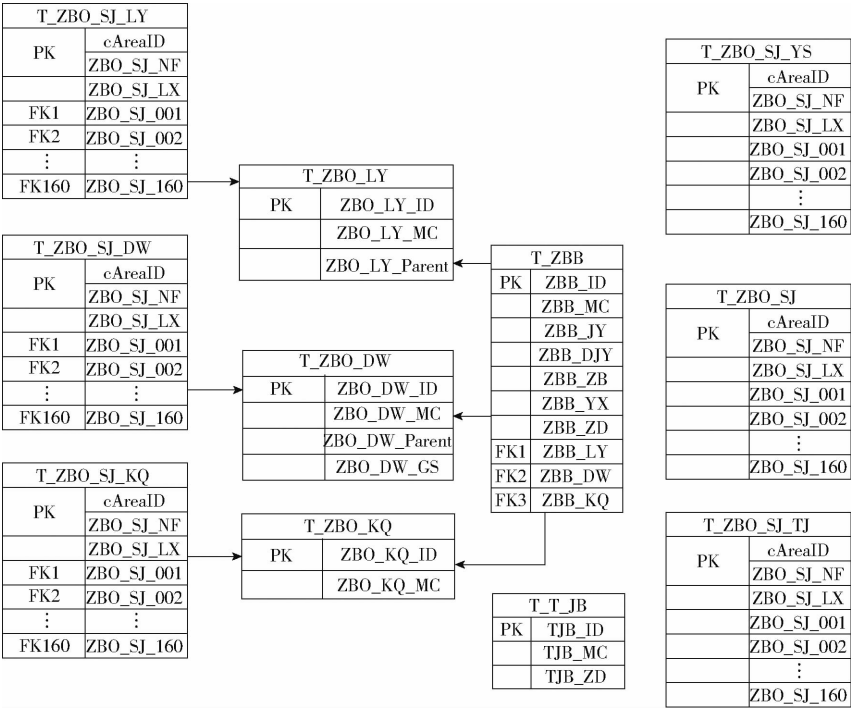


图 2 数据库表间关系图

Fig. 2 Relationship diagram between database tables

用户选择的指标数据来源单位、数值单位和数据空缺原因 3 个指标描述参数的具体选项内容。T_ZBO_SJ_LY、T_ZBO_SJ_DW、T_ZBO_SJ_KQ 和 T_ZBO_SJ_Y 这 4 个表是填报数据表,分别存放已收集的不同地区不同年份的来源单位、数值单位、数据空缺原因和数值数据。表 T_ZBO_SJ 与表 T_ZBO_SJ_Y 相对应,存放将表 T_ZBO_SJ_Y 中数据转换为标准单位后的数据。T_TJB 表是统计项结构表,是系统的辅助结构表,存放统计的内容项。T_ZBO_SJ_TJ 是统计结果表,存放填报后的统计数据。

数据校验分为单指标校验和多指标校验两部分。在结构表中,ZBB_JY 字段中存放正则表达式,作为对应指标的单指标校验条件;ZBB_DJY 字段中存放指标间的关系表达式,作为多指标校验条件;ZBB_ZB 字段用于存放指标相对应的填补方式。实现了指标与不同功能的具体实现方式间的相互对应,逻辑关系结构清晰,便于后续调用。在具体功能实现时,如图 3 所示,将具体业务处理程序中变化频繁的部分从程序中抽离,存放到结构表中,将业务处理程序转换为可动态改变的结构表和抽象控制程序^[12]。抽象控制程序通过调用结构表中的数据实现具体功能,不随具体需求的改变而变化。当需要收集的指标发生变化时,只需要增减结构表中的记录,并通过 ZBB_JY、ZBB_DJY 和 ZBB_ZB 字段调配具体指标对应的单指标校验条件、多指标校验条件和空缺数据填补方式即可,无需修改程序,极大地提高了系统的可扩展性和易维护性。

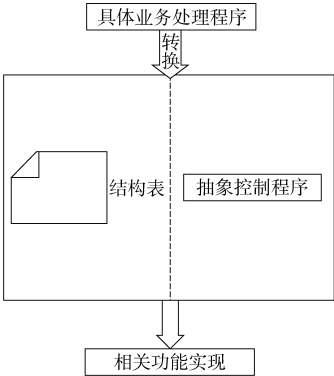


图 3 在表驱动思想下实现相关功能

Fig. 3 Realizing function through table-driven methods

受桥接模式^[20]设计思想的启发,将一个指标所对应的多个描述参数看作多个实体,每个实体对应多个选项内容。每个实体对应一张数据库表,通过增减表的个数实现描述参数的有效扩展,通过增减实体所对应的选项内容实现相应描述参数下具体选项内容的有效扩展。对于部分描述参数,提供具体选项以供用户选择,在一定程度上保障了数据录入的规范性。具体设计时,在结构表中,通过 ZBB_LY、ZBB_DW 和 ZBB_KQ 3 个字段实现指标与每个描述参数的具体选项内容间的关联。当指标的描述参数发生变化时,可增减基础数据表和结构表中相应字段;当指标描述参数中的具体选项内容发生变化时,可直接修改相应基础数据表中的具体内容。与数据校验和空缺数据填补类似,数据填报及查询界面中的指标、指标描述参数和具体选项内容均通过抽象控制程序从数据库中动态获取并显示。通过

这种可灵活扩展的设计方式,系统可很好地适应指标、指标描述参数和具体选项内容的变化。

在结构表中,ZBB_YX 字段中的内容用于控制指标的有效性。ZBB_ZD 字段的内容建立了指标与具体填报数据表和统计结果表中 ZB0_SJ_001、ZB0_SJ_002,⋯,ZB0_SJ_160 字段的对应关系,节省了大量指标名称的存储开销;同时代表相应指标的变量表示,构建了 ZBB_DJY 字段中指标间的关系表达式。同样,统计项结构表中 TJB_ZD 字段内容建立统计项与统计结果表中字段的对应关系。通过基础数据表将指标描述参数的具体选项内容编码,在存储指标数据时,以数字编码代替冗长的选项内容存储于填报数据表,实现数据压缩,从而节省传输与存储开销。基于表驱动法的可扩展数据库极大改善了系统的灵活性与普适性,为后续设计和实现 DDTMS 奠定基础。

3 动态数据上报管理系统

3.1 系统设计

本文所设计的 DDTMS 位于操作系统之上,通过网络为上层分布式应用软件——数据填报及统计系统提供数据访问、数据校验、空缺数据填补以及统计结果查询输出服务。

3.1.1 系统框架体系结构

分析数据填报及统计系统业务模式,同时考虑客户端操作方便、简单,本文设计的 DDTMS 采用 B/S3 层体系结构即 Browser/Web Server/DataBase Server。DDTMS 整个系统框架如图 4 所示。其中数据访问构件、多指标校验构件、空缺数据填补构件和统计结果查询输出构件通过 C#语言实现,运行于 Web 服务器的操作系统上;而单指标校验构件基于 jQuery 框架研发,运行于客户端操作系统上,提供单指标即时校验服务^[21]。DDTMS 通过数据访问构件与数据库交互,利用 HTTP 协议借助 Internet 与客户端浏览器进行信息通信,数据以 JSON 形式传输。

3.1.2 系统功能设计

图 5 是 DDTMS 功能模块组织结构,DDTMS 主要划分为数据库接口管理模块、数据库访问模块、单指标数据校验模块、多指标数据校验模块、空缺数据填补模块、统计结果查询模块和统计结果输出模块。数据库接口管理负责对 DDTMS 中与不同数据库交互的接口进行管理及调用;数据库访问实现对数据库中具体表及字段的操作;单指标数据校验和多指标数据校验负责对填报的指标数据进行指标及指标间的校验;空缺数据填补根据具体指标的性质对具有可填补性质的指标进行空缺数据填补;统计结果

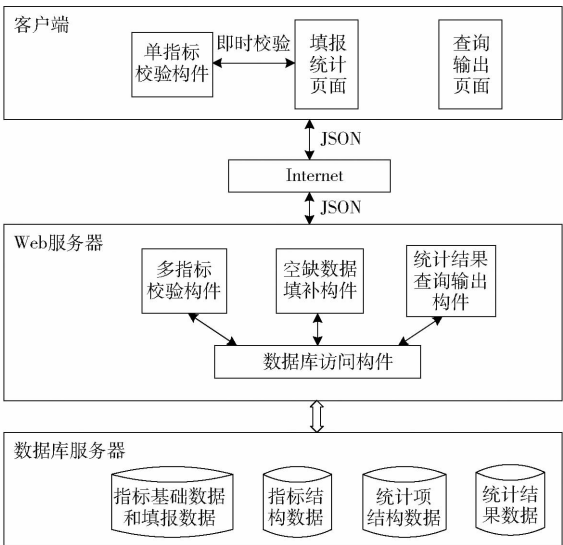


图 4 DDTMS 框架体系结构图

Fig. 4 DDTMS framework architecture

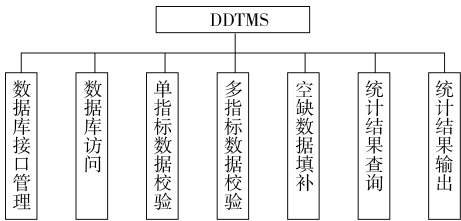


图 5 DDTMS 功能模块组织结构图

Fig. 5 Structure diagram of DDTMS functional modules

查询和统计结果输出是针对统计完成后的结果数据,实现根据用户的查询条件,从数据库中查询并输出到查询显示界面。

3.2 主要构件设计

3.2.1 数据访问构件

数据访问构件支持 DDTMS 对不同数据库访问,支撑上层业务逻辑的实现。分析其内部具体的业务实现,笔者将数据访问构件分为访问层、工厂层和 DAL 应用层。数据访问构件的设计模型如图 6 所示。

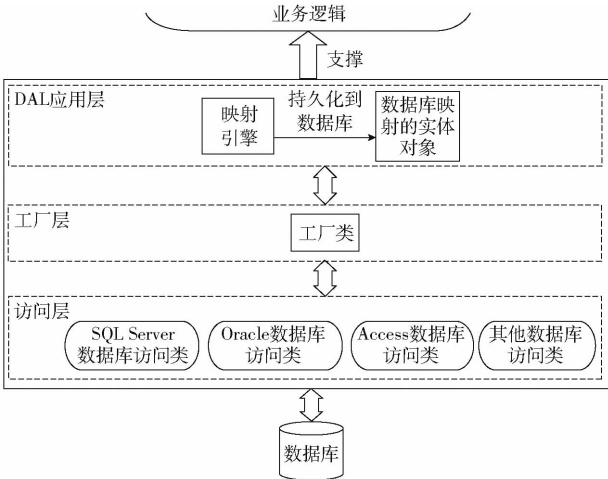


图 6 数据访问构件设计模型

Fig. 6 Design model of data access component

访问层将不同数据库中连接关闭数据库、执行 SQL 语句和执行存储过程等操作分别进行封装,形成一系列对不同数据库的访问类。支持 SQL Server、MySQL、Oracle、Access 等基本数据库。为使数据访问构件可以自动灵活地与不同数据库交互,在工厂层采用工厂模式^[22-23]与反射相结合的方式。采用工厂模式的工厂层类图如图 7 所示。SQLServerDbHelper、OracleDbHelper 和 AccessDbHelper 等为不同数据库的访问类。接口 IDbHelper 中包含一系列抽象方法以定义所需要的访问功能:连接和关闭数据库,执行一条或多条 SQL 语句、执行存储过程等等,由于包含的方法很多,在类图中以省略号

表示。每个访问类都实现接口 IDbHelper,提供各自具体的数据库访问方法。根据配置文件中选择的数据库,工厂类 DbFactory 的构造方法通过反射机制将相应数据库对应的访问类实例化,生成数据库访问对象,实现对其它数据库访问类的屏蔽。在 DAL 应用层中,为了将业务逻辑与数据库之间解耦合,以使数据库层对于业务逻辑透明,实现真正的完全面向对象,本文数据访问构件将数据库中的表映射为实体对象 Model。在对象-关系映射引擎中^[24],通过工厂类调用 GetDbHelper 方法直接获取当前数据库所对应的数据库访问对象,实现对相应实体对象 Model 的持久化。

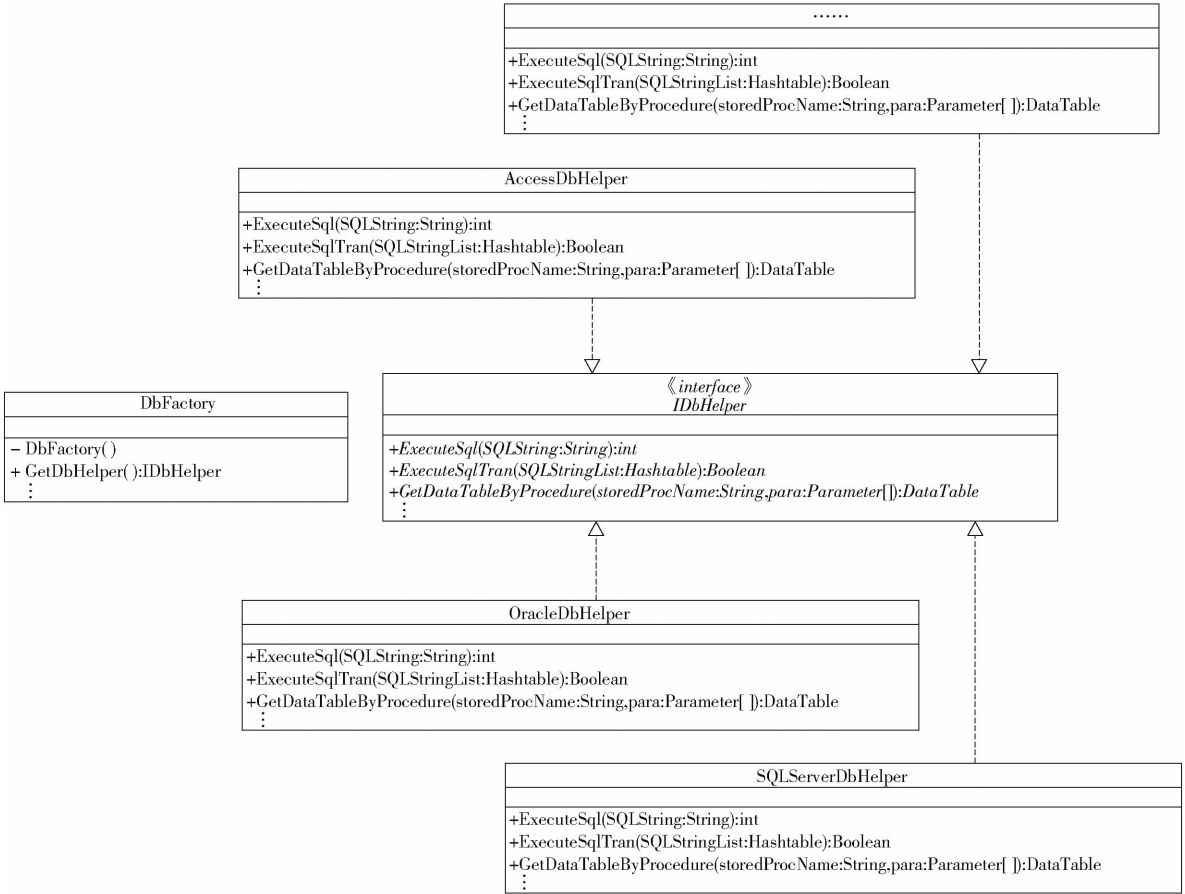


图 7 工厂层类图

Fig. 7 Class diagram of factory layer

3. 2. 2 数据校验构件

数据校验^[25-26]是指标数据填报的关键环节之一。数据填报时,多出现数值格式错误、数值空缺、数值精度不统一、数值大小不符合逻辑和指标数值间的关系不满足业务规则等情况。为了保障所收集数据的可用性,在数据录入时,对数据进行了严格的“过滤”,保障数据库中数据的准确性、可用性和完整性。为此,在 DDTMS 中设计数据校验构件,数据校验构件的架构如图 8 所示。按照检验实现的逻辑复杂度,将几种数值出现错误的情况进行划分,分别

设计单指标校验构件和多指标校验构件,单指标校验构件负责处理逻辑复杂度低的检验,而多指标校验构件负责检验较为复杂的指标间的业务逻辑。考虑到网络传输开销,单指标校验构件在客户端通过轻量级的 jQuery 方式实现,以加快检验的速度,提高系统交互性^[21]。

单指标校验构件中的单指标检验用于检验数据格式、空缺和精度,并将问题数据及错误提示返回给用户,通过用户核对修改完成校验。单指标检验模式示意图如图 9 所示。需要用户录入的指标来自于

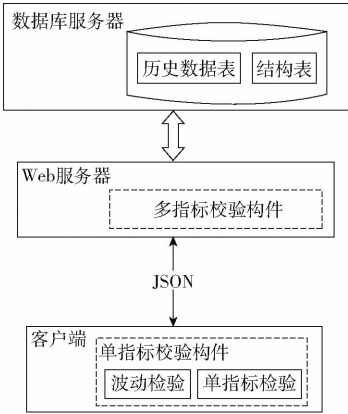


图 8 数据校验构件架构

Fig. 8 Architecture of data verification component

结构表。结构表 ZBB_JY 字段中存放相应指标数值检验的正则表达式,与用户输入的指标数值形成一一对应的关系。以此为基础,通过 jQuery 抽象控制程序实现客户端单指标检验。当指标或者具体指标的检验条件发生变化时,通过增减结构表中的记录或更改 ZBB_JY 字段中相应正则表达式式进行维护,无需更改 jQuery 抽象控制程序,方便维护的同时,可实现指标的快速扩展。针对无法严格限定其数值具体值域范围的指标,设计波动检验。通过分析数据库中的历史数据,给出预期的指标数值。将其与用户输入的指标数值比较,若相差过大,则仅仅

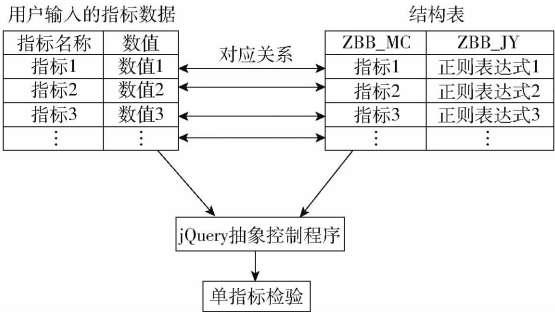


图 9 单指标检验模式

Fig. 9 Single index test

给予用户提醒,并不对指标数值进行“过滤”操作。通过波动检验,一方面规避了极端因素影响的问题,另一方面也有效解决了因用户失误导致的指标数值过大或过小问题。多指标检验时,以表达式的形式描述指标间的业务规则,其中表达式以字符串的形式存储于结构表中的 ZBB_DJY 字段。考虑多指标检验的具体实现,多指标校验构件需要在 Web 服务器完成。多指标检验设计如图 10 所示,其中 DBcheck 类中的 ZD、ZBName 和 GS 属性分别与结构表中 ZBB_ZD、ZBB_MC 和 ZBB_DJY 字段对应。图 10 中的抽象控制程序内包含了多指标检验实现的过程及方法。为了执行字符串类型的关系表达式同时判断并给出相应指标的错误信息提示,动态创建多指标检验代码,经过代码的动态编译并执行^[27],实现

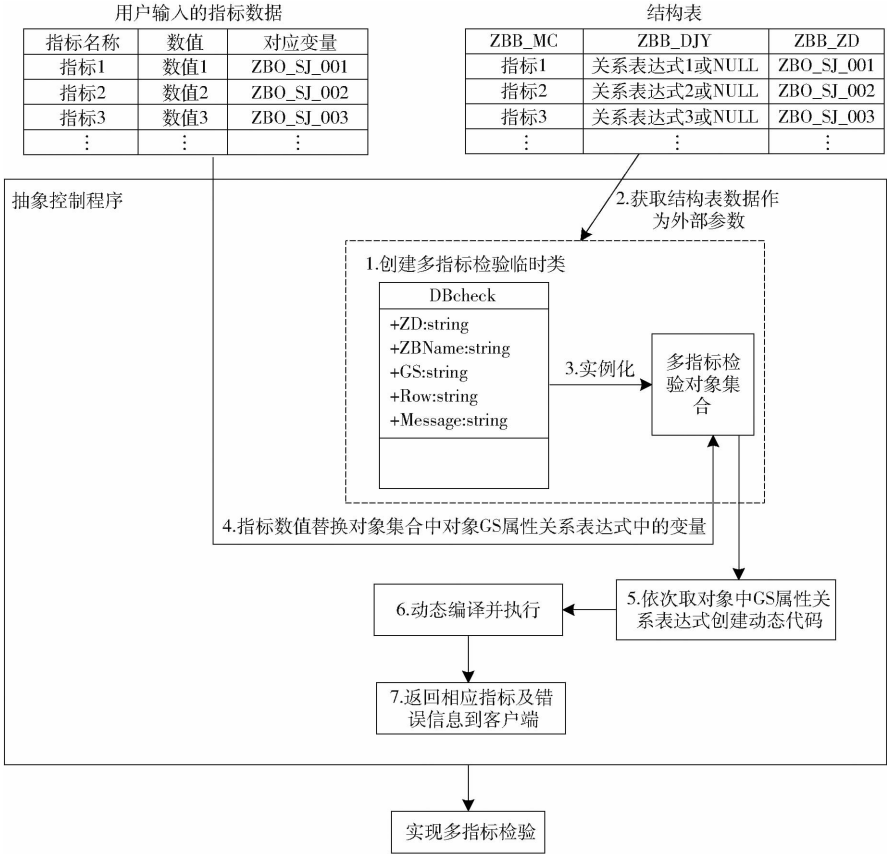


图 10 多指标检验模式

Fig. 10 Multi-index logic test

多指标检验功能。与单指标校验构件类似,当指标体系或其内部关系发生改变时,只需要更改结构表内容,无需改动抽象控制程序便可实现相应内容及多指标检验功能的动态改变。数据校验构件为可复用构件,将校验的具体需求内容与功能实现解耦合,为 DDTMS 提供数据校验功能。

3.2.3 空缺数据填补构件

由于数据采集能力有限或数据丢失等各种原因,使得用户填报时有空缺数据存在,造成系统收集的数据不全。为此,根据指标性质和具体的特性,采取合理的方式填补^[28-29]。

分析所需收集的指标性质,将可填补的指标抽象为类——可填补指标,将其中可通过相同方法填补的指标抽象为一类。可根据不同类指标实际填补需求,创建并扩展可填补指标的子类。类图如图 11 所示,人口型空间类指标、面积型空间类指标等均继承自可填补指标,作为可填补指标的子类,它们根据各自的填补算法实现了可填补指标的抽象方法:空缺数据填补方法。在明确类及类间关系的基础上,设计并完成空缺数据填补构件。其内部工作模式如图 12 所示。结构表中的 ZBB_ZB 字段给出了具体指标的填补方式,填补方式与可填补指标子类之间为一一对应关系。在空缺数据填补时,通过抽象控制程序确定具体指标的所属类,并调用相应类的空缺数据填补方法实现具体的数据填补。当指标发生变化时,只需更改结构表;当需要增添新的填补方式时,可通过创建可填补指标子类,并在子类中实现其空缺数据填补方法。将不同指标及其填补方法与填补功能的具体实现解耦合,提高了空缺数据填补构件的可扩展性,良好的可扩展性支撑其可进行指标及空缺数据填补方法的快速扩充,进而提高 DDTMS 的普适性。

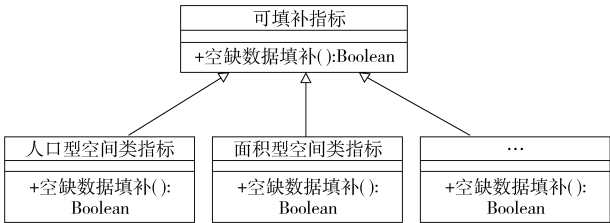


图 11 可填补指标类图
Fig. 11 Class diagram of index

3.2.4 统计结果查询输出构件

在数据填报及统计类系统中,查询并查看指标统计结果是必不可少的部分。本文总结其实现的一般过程,提取出统计结果查询输出部分,设计统计结果查询输出构件,作为解决查询查看统计结果问题的通用构件。

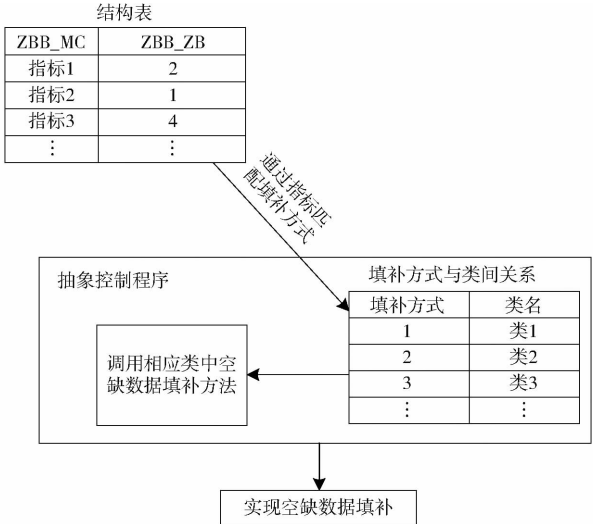


图 12 空缺数据填补构件内部工作模式
Fig. 12 Internal working model of vacancy data filling component

在统计结果查询输出构件中,将用户选择的查询条件作为输入参数,根据输入参数查询统计结果数据库,将查询出的数据封装于业务实体对象中并输出。如图 13 所示。分析用户查询条件,将输入参数分为 3 类,第 1 类为一个地区一个年份统计项查询,参数包括用户查询的地区、查询的年份和查询的统计项信息;第 2 类为多个地区一个年份统计项查询,参数包括用户查询的地区数组、查询的年份和查询的统计项信息;第 3 类为一个地区多个年份统计项查询,参数包括用户查询的年份数组、查询的地区和查询的统计项信息。抽象控制程序接收客户端传来的查询参数,根据用户查询的统计项参数,查询统计项结构表返回相应统计项及其在统计结果表中的对应字段,并将此字段作为统计项查询字段,将返回的信息装载于统计项对象集合。根据输入参数中的地区信息、年份信息以及统计项对象集合,查询统计结果表返回查询结果数据集。为接收并传输查询的结果数据,构建 ZBValue 类作为业务实体类,ZBValue 类承担数据载体的任务,用于封装并传输查询的结果数据。根据查询结果数据集的具体条数,多次实例化 ZBValue 类,生成 ZBValue 对象集合。将查询的数据封装于对象集合中,并将其输出。

4 动态数据上报管理系统的实现与测试

4.1 动态数据上报管理系统的实现

DDTMS 基于 Visual Studio 2013 平台开发,分别使用 SQL Server 2008、Oracle 和 Access 数据库实现表驱动及相关数据库设计,DDTMS 中不同构件分别采用了不同的技术和插件进行开发。

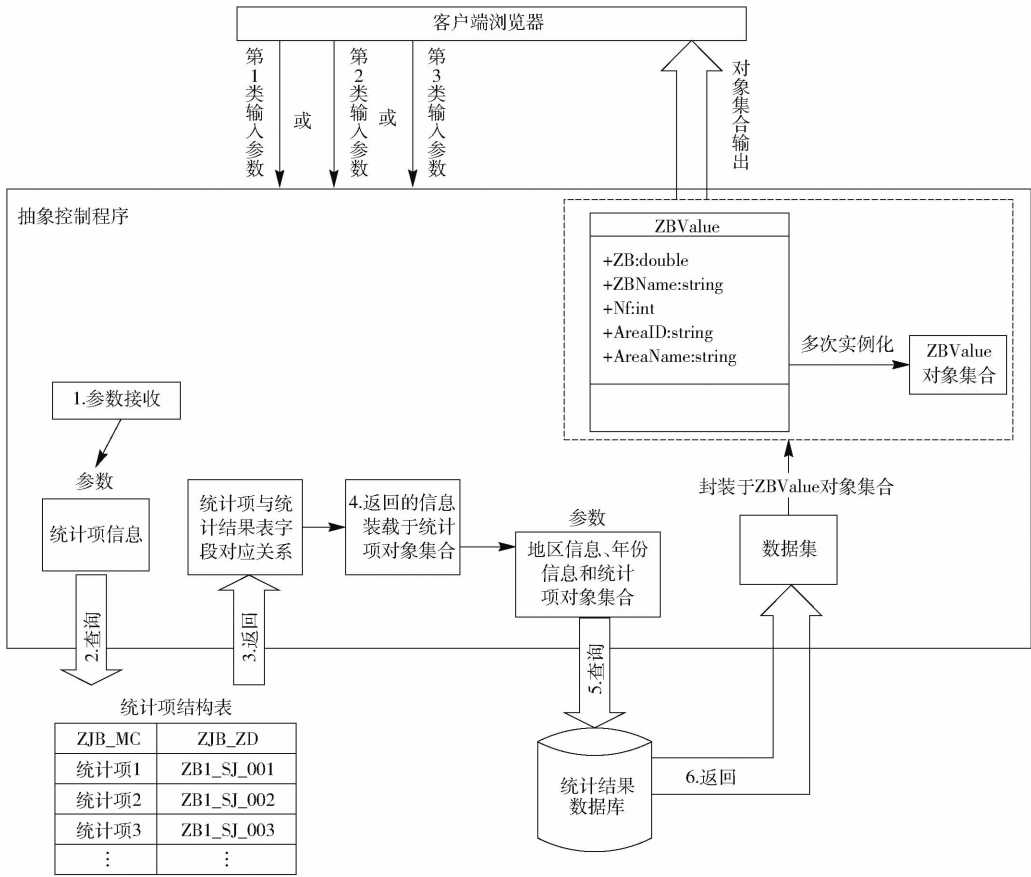


图 13 Working model of statistical result query component

4.1.1 数据访问构件的实现

数据访问构件中,为了获取更好的数据库访问性能,本文根据不同数据库采用有针对性的数据库接口,对于 SQL Server 数据库,用 SqlConnection 类连接数据库;用 SqlCommand 类执行 Transact-SQL 语句或存储过程;使用 SqlDataAdapter 类作为 DataSet 和 SQL Server 之间的桥接器。而对于 Oracle 数据库,分别采用 OracleConnection、OracleCommand 和 OracleDataAdapter,对于 Access 数据库分别采用 OleDbConnection、OleDbCommand 和 OleDbDataAdapter。除了数据库连接接口不同,不同数据库在 SQL 语法上有很多差别。为此,采用面向接口的方式解决。在接口 IDbHelper 中,定义所需要的访问功能,包括一系列重载的方法:执行 SQL 语句的方法 ExecuteSql、GetDataTable 等,执行存储过程的方法 GetDataTableByProcedure、GetDataReaderByProcedure 等,以及执行多条 SQL 语句实现事务处理的方法 ExecuteSqlTran。针对于不同数据库,分别在 SQLServerDbHelper、OracleDbHelper 和 AccessDbHelper 等类中,给出 IDbHelper 接口中方法的具体实现。在数据访问构件 DAL 应用层中,由于数据库表及字段数目众多,对数据库表进行映射时,通过开发的代码生成器进

行批量生产,生成与数据库表及表中字段具有对应关系的 Model 类:T_ZBB、T_ZBO_LY 和 T_ZB0_SJ_LY 等,以及完成相应类到数据库映射的对象-关系映射引擎:T_ZBBDAL.cs、T_ZBO_LYDAL.cs 和 T_ZB0_SJ_LYDAL.cs 等。

4.1.2 数据校验构件的实现

数据校验构件中,为改善校验时用户的体验,提高校验速度,单指标检验和波动检验均通过 jQuery 抽象控制程序实现,并分别封装于 singlecheck.js 和 fluctuationcheck.js 中。用户每输入一项数据,系统立刻校验并给出错误提示。多指标检验在服务器端实现,由于校验针对结构表中多指标校验条件进行,因此将其从数据库中取出并封装于业务实体类中,方便后续校验操作。具体实现时,创建业务实体类 DBcheck,通过对象-关系映射引擎 T_ZBBDAL.cs 操作结构表,获取结构表中记录依次装载于业务实体类 T_ZBB,选取 ZBB_DJY 字段值非空的记录,依次将数据装载于 DBcheck;同时接收客户端传来的指标及其在浏览器页面中的位置信息,依次装载于相应的 DBcheck 对象中,最终生成 DBcheck 对象集合。DBcheck 对象中 GS 属性与 Message 属性内容相同,均为使用结构表 ZBB_ZD 字段值表示的指标间的关系表达式。遍历 T_ZBB 对象集合,将

DBcheck 对象 Message 属性中内容替换为指标名称表示的关系表达式。将用户前台输入的指标数据装载于业务实体类 T_ZB0_SJ_YS。通过反射机制,获取对象 T_ZB0_SJ_YS 中与 DBcheck 对象 GS 属性中变量对应的属性值,并进行替换。之后进行一系列过程:将对象 GS 属性中的关系表达式创建为动态代码,动态编译并执行,返回关系表达式运行结果。数据校验构件把上述过程的实现封装于 DynCompilation 类的静态方法 Calc (string expression)中。遍历 DBcheck 对象集合,取对象 GS 属性值作为方法 Calc (string expression)的参数,通过 DynCompilation 类直接调用静态方法 Calc (string expression),返回关系表达式运行结果。当运行结果错误时,记录 DBcheck 对象中 Row 属性值 (Row 属性给出指标在浏览器页面所处位置)与 Message 属性值 (Message 属性给出具体错误信息提示),从而将错误信息准确地返回到浏览器页面相应位置。

4.1.3 空缺数据填补构件的实现

空缺数据填补构件中,由于用户填报的数据单位不一致,在填补空缺数据前,首先要将所有数据转换为标准单位数据。数据库 T_ZB0_DW 表的 ZB0_DW_GS 字段给出当前单位的数据转化为标准单位的数据所需要乘以的系数。遍历用户填报的数据,确定数据单位在 T_ZB0_DW 表中的 ZB0_DW_GS 值,将其与填报数据相乘得到标准单位数据。将标准单位的数据存放于 T_ZB0_SJ 表中。为保证填补数据的有效性,填补的指标数据不作为填补其他指标的依据。在填补空缺数据前,将所依据的指标数据为空的指标数据统一去除,不进行填补。根据不同类指标的性质,确定各类指标的填补算法,人口型空间类指标和面积型空间类指标均采用加权空间插值^[29]的方法。具体算法如下:

对于数据缺失年份,假设县 b 指标数据缺失且缺失数据为 W_b ,县 b 的年末总人口数为 N_b ;与县 b 同市的不同区县为县 $1, 2, \cdots, k$,与县 b 同省不同市的区县为县 $k+1, k+2, \cdots, k+h$, W_i 为县 i 的指标数据, N_i 为县 i 的年末总人口数, q 为同市临近县指标数据与人口数比所占的权重 (权重 q 的具体值在配置文件中配置)。人口型空间类指标填补算法公式为

$$W_b = \frac{\sum_{i=1}^k q N_b W_i}{\sum_{i=1}^k N_i} + \frac{\sum_{i=k+1}^{k+h} (1 - q) N_b W_i}{\sum_{i=k+1}^{k+h} N_i} \tag{1}$$

对于数据缺失年份,假设县 a 指标数据缺失且缺失数据为 V_a ,县 a 的县域国土面积为 S_a ;与县 a

同市的不同区县为县 $1, 2, \cdots, k$,与县 a 同省不同市的区县为县 $k+1, k+2, \cdots, k+h$, V_i 为县 i 的指标数据, S_i 为县 i 的县域国土面积, p 为同市临近县指标数据与面积比所占的权重 (权重 p 的具体值在配置文件中配置)。面积型空间类指标填补算法公式为

$$V_a = \frac{\sum_{i=1}^k p S_a V_i}{\sum_{i=1}^k S_i} + \frac{\sum_{i=k+1}^{k+h} (1 - p) S_a V_i}{\sum_{i=k+1}^{k+h} S_i} \tag{2}$$

4.1.4 统计结果查询输出构件的实现

在统计结果查询输出构件的抽象控制程序中,通过方法重载解决因用户查询条件不同所产生的不同类参数接收问题。查询统计结果表时,也使用方法重载,重载的方法分别为 List < ZBValue > GetZBValues (List < T_TJB > tjbs, string cAreaID, int searchdate), List < ZBValue > GetZBValues (List < T_TJB > tjbs, List < string > cAreaIDs, int searchdate) 和 List < ZBValue > GetZBValues (List < T_TJB > tjbs, string cAreaID, int [] searchdates), 方法返回 ZBValue 对象集合。以 JSON 形式将最终的对象集合向客户端浏览器输出。

4.2 动态数据上报管理系统的测试

为了验证 DDTMS 的普适性和实用性,将已通过 .NET 实现的 DDTMS 分别应用在两个不同的实际项目中,即中国林业生态资源环境承载力:数据投放器和中国生态安全指数系统。

4.2.1 中国林业生态资源环境承载力:数据投放器

系统开发时,在初期需求分析阶段,未得到确定的填报指标项,同时在有限的人力物力条件下,需要在较短的开发周期内设计并实现中国林业生态资源环境承载力:数据投放器。系统每五年进行一次指标收集工作,需要满足的基本要求:对用户填报的指标数据进行严格的校验,以保障所收集数据的准确性、可用性和完整性;为充分利用数据资源进行数据分析,对具有可填补性质的指标,采取合理的填补方式填补缺失数据;将经过统计与计算的最终评价结果供用户查询浏览。

针对上述问题,DDTMS 给出了很好的解决方案。将 DDTMS 应用到系统开发时,使用 SQL Server 2008 数据库。由于系统收集周期为 5 年,需增加描述指标数据采集时间的参数:数据年度。为此扩展数据库,增加基础数据表 T_ZB0_ND (ZB0_ND_ID, ZB0_ND_MC),同时新增 T_ZBB 表字段 ZBB_ND,作为外键以实现指标结构表与基础数据表 T_ZB0_ND 的关联。此外,只需做如下工作:配置指标结构

表、基础数据表及统计项结构表,完成简单的相关前端页面开发。配置指标结构表时,在 ZBB_JY 字段中填写用于相应指标检验的正则表达式;在 ZBB_DJY 字段中填写多指标检验的关系表达式,由于多指标检验条件涉及多个指标,只需在任一指标对应的字段中配置即可;在 ZBB_ZB 字段中填写可填补指标的填补方式。配置统计项结构表时,在 TJB_MC 字段中填写统计项名称,在 TJB_ZD 字段中,填写统计结果表中存储统计项值的字段名称。通过简单的配置,同时配合开发简单的相关前端页面,便可实现系统的基本要求,完成中国林业生态资源环境承载力:数据投放器。表驱动法及相关数据库设计支持指标及其相关功能配置的快速扩展与修改,当有新增指标或者指标体系发生变化时,只需要简单修改指标结构表和统计项结构表的内容。DDTMS 的使用极大提高了系统开发的效率和可维护性,实现了中国林业生态资源环境承载力:数据投放器的快速开发。

在实际的指标收集工作中,DDTMS 应用到中国林业生态资源环境承载力:数据投放器后,所收集数据的质量得到了保障,同时针对不同可填补指标的特性,分别使用合理的填补方法对缺失数据进行了有效的填补。单指标校验构件校验效果如图 14 所示,用户每输入一项指标数据,系统进行单指标检验和波动检验,给予用户即时信息提醒;用户填报完成所有数据时,系统集中进行多指标检验,如图 15 所示;填报工作结束后,由管理员统一进行空缺数据填补工作,填补时给予用户友好提示:“正在填补空缺数据,请耐心等待...”,效果如图 16 所示;对填补后的数据进行统计与计算,计算结果供用户查询浏览,效果如图 17 所示。



图 14 单指标校验构件效果界面

Fig. 14 Effect drawing of single index verification component

4.2.2 中国生态安全指数系统

中国生态安全指数系统所需收集的指标包含土壤有机质含量、年降水量、年平均气温和日照时数等 98 个指标,指标数目众多;同时,指标体系中出现需要通过新算法填补空缺数据的可填补指标;系统每年进行一次指标收集工作,由于不同指标全国统一



图 15 多指标检验效果界面

Fig. 15 Effect drawing of multi-index test

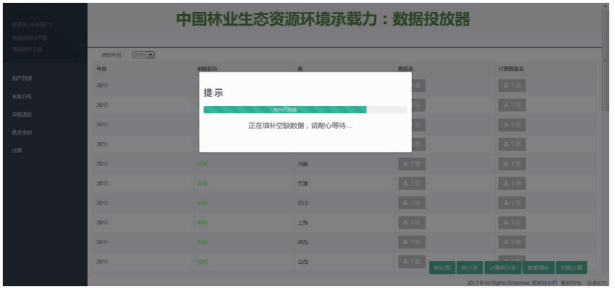


图 16 空缺数据填补效果界面

Fig. 16 Effect drawing of filling vacancy data



图 17 用户查询浏览统计与计算结果界面

Fig. 17 Results drawing of querying and browsing

采集的周期存在差异,因此不同年份需要收集的指标变化较大。系统需要满足的 3 个基本要求:①严格校验用户填报的指标数据。②对不同类别的可填补指标采取不同的填补方式进行缺失数据填补。③提供用户查询浏览统计与计算的中间结果和最终评价结果。

对于上述问题,DDTMS 也给出了很好的解决方案。将 DDTMS 应用到系统开发时,选用 Oracle 数据库。由于出现新一类可填补指标,故需要增添新的填补方式。创建可填补指标的子类:时序类指标,在填补空缺数据时,使用回归^[28-29]的方法。其具体算法如下:

假设 m 年数据空缺且缺失数据为 t_m ,取距离 m 年最近的 5 年,设最后有数据两年为 n_1 年、 n_2 年,数据分别为 t_{n1} 、 t_{n2} ,且 $n_1 < n_2$ 。时序类指标填补算法公式为

$$t_m = \begin{cases} t_{n1} \left(\frac{t_{n2}}{t_{n1}} \right)^{\frac{m-n1}{n2-n1}} & (n_1, n_2 \text{ 均存在}) \\ t_{n2} & (\text{只存在 } n_2) \end{cases} \quad (3)$$

创建时序类指标并实现父类的抽象方法空缺数据填补方法后,只需通过配置结构表中指标的 ZBB_ZB 字段值即可实现填补功能的有效扩展。由于中国生态安全指数系统的总体指标体系是确定的,供用户查询浏览的统计结果也是确定的,因此可快速配置完成指标结构表、基础数据表及统计项结构表。配置指标结构表时,将所有指标的 ZBB_YX 字段值设为 0(0 代表指标为无效指标,1 代表指标有效指标),进行指标收集工作时,通过改动指标的 ZBB_YX 字段值,来设定当前年度要收集的指标,从而解决了每年大批量转换指标所带来的一系列问题。配置统计项结构表时,将统计与计算过程中产生的统计项和最终产生的统计项均配置于统计项结构表中。之后完成简单的相关前端页面开发。在实际的项目中,DDTMS 应用到中国生态安全指数系统后,保障了所收集数据的质量,同时针对不同类别的可填补指标,采取不同的填补方法,有效填补了缺失的指标数据。用户查询浏览中间结果和最终评价结果效果如图 18 所示。

DDTMS 的使用使得系统扩展空缺数据填补功能变得简单易行,极大地提高了系统的可扩展性;对于大批量指标的变化,也只通过简单修改结构表中 ZBB_YX 字段内容即可,极大地提高了系统的可维护性;另外,DDTMS 的使用使得系统可以适应不同的数据库环境,极大提高了系统的灵活性与健壮性。



图 18 用户查询浏览效果界面

Fig. 18 Effect drawing of querying and browsing

5 结束语

阐述了数据填报及统计类信息管理系统的一般业务流程,提取此类应用的共性业务需求,设计并实现基于表驱动的动态数据上报管理系统(DDTMS)。将 DDTMS 应用到中国林业生态资源环境承载力:数据投放器和中国生态安全指数系统,以充分测试该可复用软件的实用性、稳定性和普适性。测试表明:DDTMS 缩短了应用的开发周期,极大地提高了系统开发效率;DDTMS 的使用使系统通用于不同类型的数据库,同时有效改善了系统在指标、指标描述参数等内容上以及在数据校验、空缺数据填补和统计结果查询输出等功能上的易扩展性。DDTMS 具有良好的通用性,减少了数据填报及统计类系统领域中代码冗余,并且使得此类系统的开发变得更加标准化。

参 考 文 献

1 黄焱尧. 一个通用信息管理系统模板的设计与实例化应用[D]. 北京:北京邮电大学,2013.
HUANG Qiuyao. Design of universal information management model system and the application of instantiation[D]. Beijing:Beijing University of Posts and Telecommunications,2013. (in Chinese)

2 王玲,邹小昱,刘思瑶,等. 基于 RFID 与 ZigBee 的羊场养殖信息管理系统[J/OL]. 农业机械学报,2014,45(9):247-253.
http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20140940&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2014.09.040.
WANG Ling,ZOU Xiaoyu,LIU Siyao, et al. Development of handheld terminal for sheep breeding information management based on RFID and ZigBee[J/OL]. Transactions of the Chinese Society for Agricultural Machinery,2014,45(9):247-253. (in Chinese)

3 赵国安,王晓军,刘兴森,等. 管理信息系统通用模块的设计[J]. 计算机工程,2008,34(14):49-51.
ZHAO Guoan,WANG Xiaojun,LIU Xingmiao, et al. General module design of management information systems[J]. Computer Engineering,2008,34(14):49-51. (in Chinese)

4 童小华,程效军,刘世杰,等. 基于 J2EE 高速公路网络地理信息系统平台研究[J]. 同济大学学报:自然科学版,2006,34(12):1627-1631.
TONG Xiaohua,CHENG Xiaojun,LIU Shijie, et al. Development of J2EE and WebGIS based highway management system[J]. Journal of Tongji University:Natural Science,2006,34(12):1627-1631. (in Chinese)

5 戴建国,王克如,李少昆,等. 基于国营农场的作物生产信息管理系统设计与实现[J]. 中国农业科学,2012,45(11):2159-2167.
DAI Jianguo,WANG Keru,LI Shaokun, et al. Designing and implementation of crop production management information system based on state-operated farm[J]. Scientia Agricultura Sinica,2012,45(11):2159-2167. (in Chinese)

6 何勇,余心杰. 基于 Web GIS 的农机化信息管理系统设计与实现[J]. 农业机械学报,2005,36(10):88-90.
HE Yong,YU Xinjie. Design and implementation of agricultural mechanization information management system based on Web GIS[J]. Transactions of the Chinese Society for Agricultural Machinery,2005,36(10):88-90. (in Chinese)

7 吴才聪,蔡亚平,罗梦佳,等. 基于时间窗的农机资源时空调度模型[J/OL]. 农业机械学报,2013,44(5):237-241. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20130541&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2013.05.041.
WU Caicong,CAI Yaping,LUO Mengjia, et al. Time-windows based temporal and spatial scheduling model for agricultural

- machinery resources[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(5): 237 – 241. (in Chinese)
- 8 陈智芳, 宋妮, 王景雷. 节水灌溉管理与决策支持系统[J]. 农业工程学报, 2009, 25(增刊 2): 1 – 6.
CHEN Zhifang, SONG Ni, WANG Jinglei. Water-saving irrigation management and decision support system[J]. Transactions of the CSAE, 2009, 25(Supp. 2): 1 – 6. (in Chinese)
- 9 赵玉杰, 高怀友, 师荣光. 全国基本农田环境质量监测信息填报及统计系统功能与设计[J]. 农业环境科学学报, 2004, 23(6): 1241 – 1243.
ZHAO Yujie, GAO Huaiyou, SHI Rongguang. Design and realization of agro-environmental monitoring information collection and statistic system for basic farmland[J]. Journal of Agro-Environment Science, 2004, 23(6): 1241 – 1243. (in Chinese)
- 10 ZHANG J, YONG Q, DI H, et al. A table-driven programming paradigm for context-aware application development[C]// Annual International Symposium on Applications & the Internet, 2009: 121 – 124.
- 11 YANG H C, JR D S P. Table-driven programming in SQL for enterprise information systems[C]// ICEIS 2005, Proceedings of the Seventh International Conference on Enterprise Information Systems, 2005: 424 – 427.
- 12 李订芳. 管理软件设计中的表驱动思想与软件重用[J]. 微型机与应用, 1997(3): 20 – 28.
LI Dingfang. The table driving idea and software reusing in the course of management software design[J]. Microcomputer & Its Applications, 1997(3): 20 – 28. (in Chinese)
- 13 李林, 王竹, 呼延正勇, 等. 田间数据传输同步策略与中间件研究[J]. 农业机械学报, 2016, 47(1): 279 – 288. http://www.jcsam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20160138&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2016.01.038.
LI Lin, WANG Zhu, HUYAN Zhengyong, et al. Research on field data transmission synchronization strategy and middleware[J]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(1): 279 – 288. (in Chinese)
- 14 王孝明, 胡健, 陆坤, 等. 基于 .NET 平台可复用软件框架的设计与实现[J]. 计算机工程, 2004, 30(22): 76 – 78.
WANG Xiaoming, HU Jian, LU Kun, et al. Design and implementation of reusable software framework based on .NET[J]. Computer Engineering, 2004, 30(22): 76 – 78. (in Chinese)
- 15 YAO H, ETZKORN L H, VIRANI S. Automated classification and retrieval of reusable software components[J]. Journal of the Association for Information Science and Technology, 2008, 59(4): 613 – 627.
- 16 SCHMIDT D C, BUSCHMANN F. Patterns, frameworks, and middleware: their synergistic relationships[C]// 2003 International Conference on Software Engineering, 2003: 694 – 704.
- 17 吴泉源. 网络计算中间件[J]. 软件学报, 2013, 24(1): 67 – 76.
WU Quanyuan. Network computing middleware[J]. Journal of Software, 2013, 24(1): 67 – 76. (in Chinese)
- 18 王珊, 萨师煊. 数据库系统概论[M]. 北京: 高等教育出版社, 2006: 172 – 182.
- 19 刘洋, 高连生, 王斌. 一种面向应用扩展的树状数据库设计模型[J]. 计算机工程与设计, 2006, 27(21): 4074 – 4077.
LIU Yang, GAO Liansheng, WANG Bin. Tree-like database design model for expansible application[J]. Computer Engineering and Design, 2006, 27(21): 4074 – 4077. (in Chinese)
- 20 QIU W, ZOU W, SUN Y. Design patterns applied in power system analysis software package[C]// 2012 International Conference on Industrial Control and Electronics Engineering, 2012: 836 – 840.
- 21 DHAND R. Reducing web page post backs through jQuery ajax call in a trust based framework[C]// 2012 International Conference on Computing Sciences, 2012: 217 – 219.
- 22 柴晟, 李明富, 罗莉娟, 等. 基于设计模式构建数据访问中间件[J]. 计算机工程与设计, 2007, 28(17): 4102 – 4104.
CHAI Sheng, LI Mingfu, LUO Lijuan, et al. Building data access middleware based on design pattern[J]. Computer Engineering and Design, 2007, 28(17): 4102 – 4104. (in Chinese)
- 23 王彬, 靳大尉, 郝文宁, 等. 设计模式在数据库访问权限系统中的应用[J]. 计算机应用, 2012, 32(增刊 2): 113 – 115.
WANG Bin, JIN Dawei, HAO Wenning, et al. Application of design patterns in database access permission system[J]. Journal of Computer Applications, 2012, 32(Supp. 2): 113 – 115. (in Chinese)
- 24 周源. 对象关系映射引擎的设计与实现[D]. 南京: 东南大学, 2005.
ZHOU Yuan. Object relational mapping engine design and implementation[D]. Nanjing: Southeast University, 2005. (in Chinese)
- 25 LING Y, AN Y, LIU M, et al. An error detecting and tagging framework for reducing data entry errors in electronic medical records (EMR) system[C]// IEEE International Conference on Bioinformatics and Biomedicine, 2013: 249 – 254.
- 26 刘成, 张凯, 陈建勋. 混合方式数据验证方案的研究[J]. 计算机工程与设计, 2013, 34(1): 366 – 371.
LIU Cheng, ZHANG Kai, CHEN Jianxun. Mixed-mode data validation program research[J]. Computer Engineering and Design, 2013, 34(1): 366 – 371. (in Chinese)
- 27 廖柯. 基于动态编译的泵业电子商务平台设计与实现[D]. 上海: 复旦大学, 2011.
LIAO Ke. Design and implementation of pump e-commerce platform based on dynamic compilation[D]. Shanghai: Fudan University, 2011. (in Chinese)
- 28 武森, 冯小东, 单志广. 基于不完备数据聚类的缺失数据填补方法[J]. 计算机学报, 2012, 35(8): 1726 – 1738.
WU Sen, FENG Xiaodong, SHAN Zhiguang. Missing data imputation approach based on incomplete data clustering[J]. Chinese Journal of Computers, 2012, 35(8): 1726 – 1738. (in Chinese)
- 29 刘爱利, 王培法, 丁圆圆. 地统计学概论[M]. 北京: 科学出版社, 2012.