

# 基于XML的信息工程监理异构数据集成研究

郎星建, 刘宏志

(北京工商大学 计算机与信息工程学院, 北京 100048)

**摘要:**针对信息工程监理数据管理的特点,采用XML技术与Web Services技术开发出了一个三层架构的信息工程监理异构数据集成系统。该系统结合了XML技术与全局数据模式,以XML技术进行数据的交互,转换来自不同数据源的数据,对各异构数据源中的数据进行集成,为各异构数据源建立可共享的集成查询平台。应用结果表明,该异构数据集成系统具有良好的透明性、各平台独立性以及可共享性,较好地解决了信息工程监理中异构数据的管理。

**关键词:**信息工程监理;异构数据;可扩展标记语言;Web服务

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2011)03-0221-04

## Research on Heterogeneous Data Integration for Information Engineering Surveillance Based on XML

LANG Xing-jian, LIU Hong-zhi

(College of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China)

**Abstract:**Based on the features of data involved in information engineering surveillance, a data integration system for information engineering surveillance with three layers architecture was devised by XML technology and Web Services technology. The data integration system combined XML technology with overall data model. Based on data interacting by XML technology in middle layer, the system can converse and integrate all types of heterogeneous data from different data sources, so that it can provide heterogeneous specimen-databases with unified inquiry platform. The application results show that the system was characterized by transparency and independence and sharing, so that it benefited greatly to data management for surveillance of information system engineering.

**Key words:**information engineering surveillance; heterogeneous data; XML; Web services

### 1 概述

依据信息产业部《信息系统工程监理暂行规定》,信息工程监理是指依法设立且具备相应资质的信息系统工程监理单位,受业主单位委托,依据国家有关法律法规、技术标准和信息系统工程监理合同,对信息工程项目的实施进行监督管理。

监理的主要内容是对信息工程的质量、进度和投资进行监督,对项目合同和文档资料进行管理,协调有关单位间的工作关系。由此可见,信息工程监理过程实质上是信息管理的过程,即监理单位(监理工程师)受业主的委托,在明确监理信息流程的基础上,通过建立一定的组织机构,对信息工程监理信息进行收集、加工、存储、传递、分析和应用的过程。更重要的

是,信息管理是信息工程监理“四控、三管、一协调”<sup>[1]</sup>中的其中一项,对控制信息工程监理中三大目标(质量控制、进度控制和投资控制)的实施起着重要的作用。

但是,监理过程中形成的信息具有以下特点:

(1)数据分散分布。数据不但位于监理方,还位于工程的业主方、工程的各承建方、工程的各承包方、商品的供应方等,如图1所示。

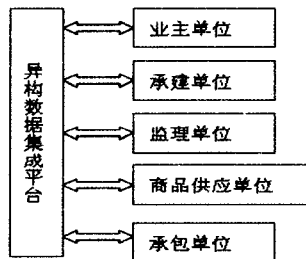


图1 信息工程监理异构数据分布图

(2)数据量巨大。按工程建设信息的用途划分可分为投资控制信息、进度控制信息、质量控制信息、合

收稿日期:2010-06-27;修回日期:2010-09-28

基金项目:北京市科技发展计划面上项目(KM200610011008)

作者简介:郎星建(1978-),男,硕士,研究方向为计算机网络与安全;刘宏志,博士,教授,研究方向为信息工程监理、电子政务、软件工程、计算机网络。

同管理信息、组织协调信息、法律法规信息等。

(3) 数据异构。该异构性又分为平台异构性、数据库系统异构性和语义异构性<sup>[2]</sup>。在以往的信息化工程建设监理中,由于缺少有效的管理和沟通的桥梁、工具,导致信息工程建设各方信息失真,沟通困难,严重影响了信息工程监理三大目标(质量控制、进度控制和投资控制)。为了提升工程建设质量和加快建设速度,对这些异构数据进行集成非常有必要。

文中基于 XML 和 Web Services 技术探讨了信息工程监理中异构数据的集成,解决了数据的异构性,提供了数据访问的透明性,保证了各单位数据源的自治性。

## 2 异构数据集成技术

### 2.1 XML 技术

XML 语言是为文档交换所设计,以一种开放的、自我描述方式定义的数据结构,即描述数据更突出结构的描述,实现了跨平台数据交换和数据传输,已经成为事实上的数据交换标准<sup>[3]</sup>。

统一的异构数据库系统的信息交换工具,必须有一个公共数据模型,将关系模型映射到该公共数据模型,再将公共模型映射回关系模型<sup>[4]</sup>。由于 XML 的纯文本与平台无关性首先满足了解决异构关系数据源所需要的跨平台性;同时 XML 强大的结构性和良好的语义性满足了表达关系数据库的结构和各种约束的需要;XML 优良的交互性为转换带来了方便,使得数据易于操纵;XML 的易于扩展性使得应用可以进一步扩展;XML 的可格式化让转换出的结果有更多的表现形式。并且由于 XML 已成为事实上的数据交换标准,因此,XML 完全可以作为异构数据库转换的公共数据模型。

### 2.2 Web Services 技术

Web 服务是一种在 Internet 环境下的松散耦合的 Web 应用之间的相互调用、相互集成的技术框架<sup>[5]</sup>。Web 服务像组件一样,用户可以重用它而不用关心它是如何实现的。它的体系构架是一种新的体系构架,为面向服务的构架 SOAX。Web 服务提供了一组用于提供服务的接口,用户可以方便地通过这些接口调用 Web 服务提供的相应功能。Web 服务使用 HTTP 和 XML 等标准协议和数据格式进行通信。

使用 SOAP 协议统一数据交换格式可保证信息工程信息在异构平台之间的无障传输。SOAP 协议可以穿过任何防火墙,并且 SOAP 数据包中包含着以 XML 编码的数据,易于分析和使用<sup>[6]</sup>,此外 SOAP 具有很好的伸缩性,能同时为非常多的用户服务。

WSDL 文件用于说明消息格式的表示法,以 XML

架构标准为基础,这意味着它与编程语言无关而且以标准为基础,因此适用于说明从不同平台以不同编程语言访问的 Web Service 接口,此外 WSDL 还定义了服务的位置以及使用什么通信协议与服务进行通信。

Web 服务架构由三种角色组成:服务提供者、服务请求者、服务注册中心<sup>[7]</sup>。这三个角色通过三个基本操作:发布、查找、绑定来互相作用。服务提供者向服务注册中心发布服务。服务请求者通过服务注册中心查找所申请的服务,并绑定到这些服务上。

因此 Web Services 是一种动态的集成解决方案,使系统具有高度的灵活性和伸缩性,克服了过去使用的基于 RPC 和 API 的集成技术的缺陷,这也满足了松散耦合的要求。

## 3 异构数据集成系统设计目标

异构数据集成的目标是整合多个异构数据源,向用户提供统一查询更新等操作界面,使得用户透明地操作多个异构数据源。异构数据源是指由于数据源所处的平台异构、数据格式上的异构、数据语义上的不同而产生的异构<sup>[8]</sup>。

异构数据集成系统设计的主要目标:按照统一标准规范,建立多层次、多用途、兼容性好、功能性强的分布式数据库系统,为信息监理各方提供强大的数据支持与服务,实现对各类信息的查询、检索、统计、分析等具体服务,提高信息工程建设水平。为此要着重解决好以下问题:数据的各种异构性问题,各异构系统的自治性问题,信息工程监理各方对整个异构系统的透明性操作问题,以及系统的扩展升级问题。

文中设计的信息工程异构数据集成系统结构,是在对各种异构数据集成方法的分析研究的基础上,采用 XML 与 Web Services 技术构建。设计目标是将全局数据模式和 XML 技术相结合,以 XML 格式的文档作为中介进行数据的交互,各异构数据源的数据能够相互转换,在此基础上建立起各异构数据源统一的查询平台,方便信息工程各方即时地对数据进行查询与操纵。另外,应确保该系统降低耦合度,保持自治性,提高重用度,增强扩展性。

## 4 系统设计方案

该信息工程监理异构数据集成系统以异构数据源管理系统为基础,目标是对工程各方的异构数据源数据进行集成,协调数据源之间的通信,以方便工程各方对集成数据的共享与访问。数据层为系统中各个异构数据源;位于中间层的是异构数据集成处理中间件,进行业务逻辑的处理;位于应用层是应用程序。该系统的结构图如图 2 所示。

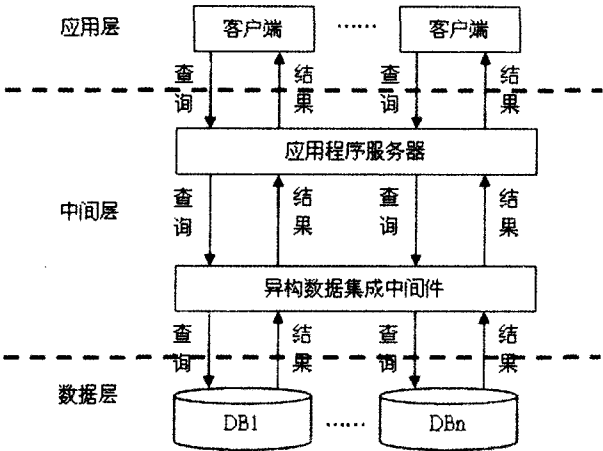


图2 异构数据集成三层模型

查询时,应用层发出查询请求,中间层接收客户端发送的查询请求。该查询可能需要访问数据层中不同数据源中的信息,所以该请求为全局查询请求。异构数据集成中间层接收该全局查询,然后分解成对各具体数据源的子查询。服务定位器根据子查询涉及到的数据源,定位具体数据源,对数据层的具体数据源发出查询,并把子查询结果返回到中间层,中间层整合各子查询结果,将整合后的数据作为从应用层的全局查询的结果返回,从而完成了一次查询过程。

在上述查询过程中,因为中间层屏蔽了对各后台数据源的操作<sup>[9]</sup>,所以对整个异构数据系统的查询相对于用户来说查询过程都是透明的。

5 系统详细设计

依据对信息工程监理系统的分析,文中设计了一种利用XML格式文档做为数据交换格式,利用XML Schema建立公共模型,采用中间件方式集成模式的数据集成系统的实现方案。该异构数据集成系统由三层组成:数据层,中间层,应用层,如图3所示。

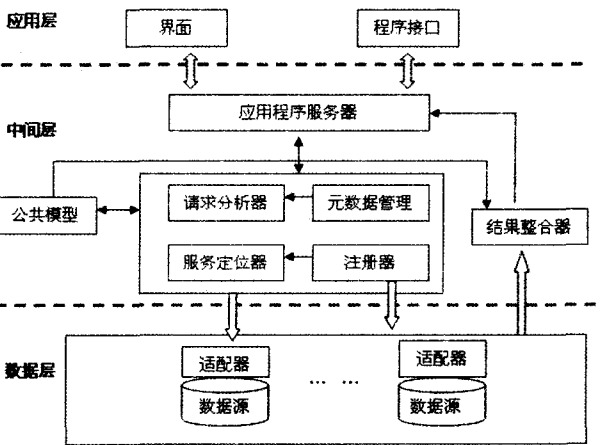


图3 信息工程监理异构数据集成平台架构

5.1 应用层

应用层包括用户操作界面以及供其他应用程序可调用的接口。此层可接受用户的操作与其它应用程序的调用,并接收中间层的操作结果,把结果呈现给用户和返回给调用接口的程序。由于用户只和应用层交互,而底层对各异构数据的操作对用户来说是透明的,整个查询就像在一个数据源里进行操作一样。

5.2 数据层

数据层是整个系统的基础。该层由各个不同的数据源组成,不同的数据源为整个集成系统提供服务,并将其自身的服务注册到注册器中。每个数据源都有一个适配器。适配器由查询转换器、Web服务和结果转换器构成,该异构数据集成系统中每个异构数据源都有自己独立的适配器。适配器的作用就是把该数据源提供的操作描述成服务,生成服务描述文档,对外提供一个统一的查询请求接口。通过适配,对数据源的调用转换成了对包装器中的Web服务的调用,从而确保各数据源的自治性。查询数据时XQuery子查询通过查询转换器映射为对某个数据源的实际物理查询。输出数据时结果转换器把子查询的数据结果转化成统一的XML数据格式,并把结果返回给上一层。

5.3 中间层

中间层是整个异构集成系统的业务逻辑部分,是整个系统的核心层。中间层接受应用层的请求,并判断请求类型。如果是查询请求,则根据元数据管理器将查询分解成若干个针对各个异构数据源的子查询。中间层的服务定位器定位查询涉及到的异构数据源,然后对涉及到的数据源查询并返回到该层,结果整合器对各子查询返回的数据整合得到查询结果,回送到应用层。如果是更新请求,就由服务定位器定位到相关的数据源,调用服务,更新数据源。中间层中主要模块的功能描述如下:

(1)公共模型。

公共模型包括异构集成系统的XML Schema全局模式,各局部数据源的XML Schema输出模式,以及全局模式与各输出模式之间的映射关系等。由于XML良好的跨平台传输,可以将上面的数据存储成XML格式的文档。公共模型为全局查询的分解和查询子结果的整合提供了参考依据,所以公共模型是整个异构集成系统的运行基础。

集成多个异构数据源时,提供给用户一个全局概念集合,此集合集成了整个系统数据。根据这个全局概念集合用户产生查询请求时,不用再考虑各局部数据源语义不一致问题<sup>[10]</sup>,从而可以做到透明查询整个集成系统。整个集成系统的数据信息全局视图由全局模式定义。正是根据全局模式所定义的全局共享概

念,用户提交全局查询。各具体数据源的概念和关系是由局部模式来定义的。每一个局部模式都对应到全局模式上,对应的结果就是全局模式与各数据源的局部模式间的映射关系。用户提交全局查询时,系统根据这种映射关系将查询重构为对各具体数据源的子查询,返回查询结果时,系统也依据这种映射关系将各子查询结果整合,用户最终得到所需的集成查询结果。

### (2) 注册器。

一个数据源把自己的操作描述成服务以后,就需要把服务的相关信息放置在注册器中,即注册服务。系统根据数据源在注册器中提供的服务册,向服务定位器提供查询服务,并在服务调用向系统提供服务的详细信息。

### (3) 请求分析器。

请求分析器是本集成系统的重要组成部分。它可以把用户请求转换成针对各个数据源的子请求。应用层发送请求到请求分析器,请求分析器分析请求。若为查询请求,则分解成针对各个数据源的子查询语句。若为更新请求,则定位需要更新的数据源并进行更新操作。查询处理包括查询分解、查询转换和全局优化三部分<sup>[11]</sup>。全局查询算法如下:

#### [算法 1] 解析全局查询

输入: 查询语句  $Q$

输出: 一个或若干个针对一个数据库的子查询

$Q_1, Q_2, Q_3, \dots, Q_n$

算法: for 共享模型数据库的个数

建立子查询( $Q_1, Q_2, Q_3, \dots, Q_n$ )

End for

[算法 2] 分解  $Q_1, Q_2, Q_3, \dots, Q_n$  的 where 子句

输入:  $Q_1, Q_2, Q_3, \dots, Q_n$

输出: 多个子查询( $Q_{11}, Q_{12}, \dots, Q_{1m}, Q_{21}, Q_{22}, \dots, Q_{2m}, Q_{n1}, Q_{n2}, \dots, Q_{nm}$ )

算法:

for  $Q_1, Q_2, Q_3, \dots, Q_n$  的个数

for  $Q_n$  的 where 子句中的所有查询条件

建立子查询( $Q_{n1}, Q_{n2}, Q_{n3}, \dots, Q_{nm}$ )

end for

end for

### (4) 结果整合器。

结果整合器把从各个数据源查询的结果转换成 XML 数据进行汇总整理,并参考公共模型进行集成处理,然后将集成结果发送到应用程序服务器,然后以 XML 数据格式返回应用层。

### (5) 服务定位器。

服务定位器的作用是接受请求分析器的查询或更

新请求,然后对数据源前端的适配器进行服务定位,如果查找到服务,则对数据源发出服务请求。如果没有找到就发送失败信息到结果整合器。

(6) 元数据管理器。所谓元数据就是用来描述数据资源的数据<sup>[12]</sup>,是信息共享和交换的基础和前提,它为整个系统提供了一种数据标准,及各个数据源与统一元数据标准的映射,它具有以下作用:

1) 通过元数据用户能够了解数据的来源、数据采集时间、数据所跨的空间范围、数据精度、数据质量、数据的可用性等。

2) 通过元数据用户能够了解数据的存储介质、数据的所有权归属、数据的作者信息和联系人信息以及数据的获取方法和途径。

3) 通过元数据用户能够了解数据的存储格式、数据量等信息。

总之,元数据应当提供用户了解数据的一切必要信息并有利于数据集的查询检索。

整个系统的工作流程如图 4 所示。

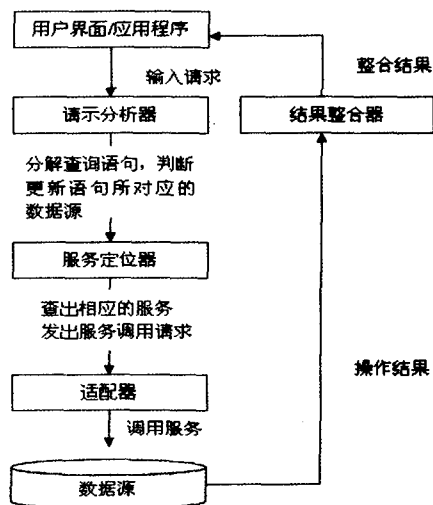


图 4 信息工程监理异构数据集成工作流程

(1) 用户通过界面提出操作请求,并将结果传送给请求分析器。

(2) 请求分析器若收到查询请求,则将查询分解成对应数据源的查询请求。若收到更新请求,则判断对数据源做哪些更新操作。

(3) 服务定位器根据请求分析器传来的查询语句与服务注册中心的服务进行匹配,找到相应的服务并向适配器发出调用请求。

(4) 适配器收到服务请求,激活服务,并把结果返回给结果整合器。

(5) 结果整合器把各个数据源传来的操作结果整合,并把结果返回给界面。

(下转第 228 页)

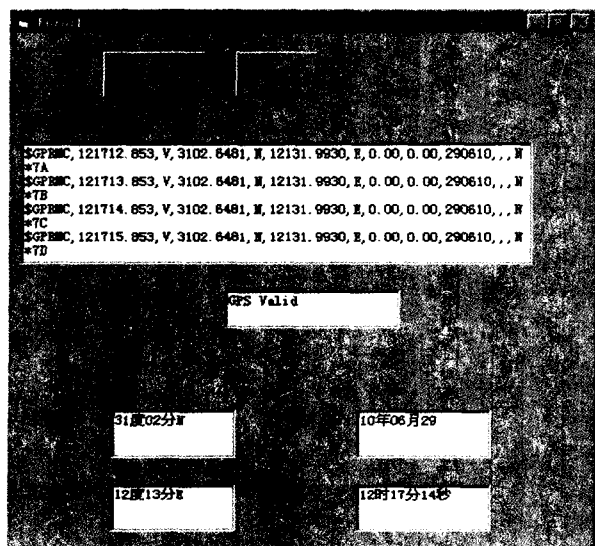


图 3 运行结果

### 3 结束语

文中介绍了如何利用 VB 的 MSComm 控件对 GPS 定位信息进行提取,并将提取的定位信息通过 GPS 接收机和 PC 机之间的数据通信实时地显示在计算机上,达到了监测公路路面、提高路面的平整度、实现列车安全行驶的目的。这项技术为路面平整度的动态检测提供了必要的技术支持,具有一定的实用价值。

#### 参考文献:

- [1] 盛安连,戴经樑.高等级公路路面平整度传递理论及其应

(上接第 224 页)

(6) 用户界面把操作结果显示给用户。至此完成了对异构数据的查询。

### 6 结束语

通过综合利用 XML 技术、Web Services 技术及 Java 技术,设计出了信息工程异构数据集成平台,经实际应用验证,达到了预期的共享性、透明性、自治性、可扩展性目标,为信息工程监理的异构数据管理发挥了极大的便利性。

#### 参考文献:

- [1] 刘宏志,葛迺康.信息化工程监理[M].北京:中国电力出版社,2009.  
[2] 李文正.电子政务与城市应急管理[M].北京:中国水利水电出版社,2008.  
[3] 史哗翎,黎建辉.关系数据库模式到 XML Schema 的通用映射模型[J].计算机工程,2009,35(7):35-38.  
[4] 袁梅冷,黄烟波,黄家林,等. J2EE 应用模型中 MVC 软件

用[J].中国公路学报,1992,5(4):8-16.

- [2] 吴景海,王德群,商耀祥.高速公路路堤和路面软基病害检测及加固处理[J].中国公路学报,2003,16(2):13-17.  
[3] 周 键.基于 GPS/GIS/GSM 的公路隧道运营管理系统研究[J].中国公路学报,2004,17(3):112-116.  
[4] 陈欣林.车载 GPS 接收机与 PC 机间串行通讯的实现[J].技术与创新管理,2009,30(1):99-101.  
[5] 胡 静.Windows 环境下计算机与 GPS 串行通信程序的开发[J].计算机应用研究,2002(3):89-93.  
[6] 李华贵,项志华,何 伟,等.基于 GPS 和 GPRS 车载导航定位系统的实现[J].计算机技术与发展,2006,16(11):241-245.  
[7] 王 忠,许凡草,任苏萍.多功能接口 GPS 接收系统研究[J].电子测量技术,2007(12):112-115.  
[8] Liao L, Patterson D J, Fox D, et al. Building Personal Maps from GPS Data[M]. [s. l.]: Springer Press, 2005: 249-265.  
[9] Ashbrook D, Starmer T. Using GPS to learn significant locations and predict movement across multiple users[J]. Personal and Ubiquitous Computing, 2003, 7(5): 275-286.  
[10] 何香玲,郑 刚. GPS 通信的 NEMA 协议及定位数据的提取[J].计算机应用软件,2001,21(12):121-122.  
[11] Zheng Y, Liu L, Wang L, et al. Learning transportation mode from raw GPS data for geographic applications on the Web [C]// In: Proc. WWW 2008. [s. l.]: ACM Press, 2008: 247-256.  
[12] 于志刚,卢秀山,张莉娜,等. NMEA-0183 协议及基于 VB 数据提取的实现[J].山东科技大学学报:自然科学版,2006,25(4):82-86.

体系结构的研究与应用[J].计算机应用研究,2003(3):88-89.

- [5] 赵 艳,李 钧.异构数据源分布式事务处理研究[J].计算机工程,2009,35(4):69-71.  
[6] 邹晓静,刘 伟,卢贤玲.基于 XML 的异构数据库系统的研究[J].微计算机信息,2007(8):160-162.  
[7] 邵 敏,李力鸿.XML 编程实践[M].北京:清华大学出版社,2002:49-71.  
[8] 陶以政.基于 Java 和 XML 技术的异构信息系统数据集成框架应用研究[J].计算机应用研究,2004(5):38-40.  
[9] 王 理,陈 皓,夏 辉,等.在异构数据库环境中实现数据集成[J].现代电子技术,2006(6):83-84.  
[10] Tansalarak N, Claypool K T. QMatch - Using paths to match XML Schema[J]. Data & Knowledge Engineering, 2007, 60: 260-282.  
[11] Madria S, Passi K, Bhowmick S. An XML Schema integration and query mechanism system[J]. Data & Knowledge Engineering, 2007, 65: 266-303.  
[12] Ronald B. XML and Database[EB/OL]. 2003-02-01[2007-10-12]. http://www.rpbouret.com/xml-d-bms.