

# XML在动车段(所)异构数据源集成中的应用

吴静<sup>1</sup>, 张春<sup>2</sup>, 张杰<sup>2</sup>, 杨留君<sup>2</sup>(1. 北京交通大学 软件学院, 北京 100044;  
2. 北京交通大学 计算机信息与技术学院, 北京 100044)

**摘要:** 动车段(所)信息化是铁路信息化的一个重要组成部分, 动车段(所)每个部门的信息就是一个数据源, 由于各部门建立时间不同, 导致了这些数据源均为异构、分布的, 从而构成了复杂庞大的数据源环境, 形成了一个个“信息孤岛”。因此非常需要一种信息技术来实现动车段(所)各系统的应用集成及信息共享, 并为决策提供及时可靠的服务。文中介绍了基于XML的中间件技术在动车段(所)信息集成中的应用, 通过对每一个异构数据源增加一个以XML为中间格式的封装体, 即在不改变数据源的前提下, 用XML对数据源的相关信息进行了封装, 用以实现动车段(所)管理信息系统内部各子系统间, 以及该系统与外部系统之间的数据交换和访问。

**关键词:** XML; 中间件; 信息集成; 中介器; 包装器

**中图分类号:** TP393.07

**文献标识码:** A

**文章编号:** 1673-629X(2006)12-0148-02

## Application of XML in Information Integration and Sharing of Rolling Stock Depot

WU Jing<sup>1</sup>, ZHANG Chun<sup>2</sup>, ZHANG Jie<sup>2</sup>, Yang Liu-jun<sup>2</sup>(1. School of Software, Beijing Jiaotong University, Beijing 100044, China;  
2. School of Computer Information and Technology, Beijing Jiaotong University, Beijing 100044, China)

**Abstract:** The rolling stock depot information system is one of the most important part of railway information system. Every information point of rolling stock depot is a data pool. Those data pools are separated and are in different structure because of the different time when these departments were set up. Those make the DBS environment stricter, larger and more complex, and form several separated "Information islands". So need a kind of technology that can implement the information integration and sharing of different sub-systems, and can provide creditable services for the management and decision-making. Introduce a kind of information integration technology that is based on XML and mediated schema. Though this method, adds a sealed body which use XML as its middleware-format to each isomerism data pool. That means under the premise of don't changing any of the data pool, use XML to seal the related information of them, in order to implements the data exchange among the interior various subsystems of the rolling stock depot information system, and also that between the rolling stock depot information system and external system.

**Key words:** XML; middleware; information integration and sharing; mediator; wrapper

### 0 概述

目前, 中国铁路信息化建设已取得了丰硕的成果。运输管理信息系统 TMIS 已部分投入运用, 铁路客票发售和预订系统(PMIS)基本建成投入运用, 而调度管理信息系统(DMI)S、车辆管理信息系统(CMIS)、财务管理信息系统(RFAMIS)和全国铁路信息服务系统等也已相继开发。作为铁路信息化重要组成部分的动车段(所)信息化建设也在紧锣密鼓地进行中。但是由于目前行业内缺乏统一的规范和标准, 造成了信息系统的重复建设, 并且各系统

之间以及各系统内部的互联性、互通性和互操作性很差, 分散的异构数据直接导致信息需求的不足, 难以为其他系统及其管理人员及时利用。针对上述问题, 解决的方法是采取 XML 技术作为不同数据源的中间数据格式, 将来自不同数据源的数据统一到 XML 存储库中, 进行转化和加工, 以实现大规模的异构、分布数据的共享和利用。

### 1 XML 技术简介

XML(Extensible Markup Language)是由 W3C 组织于 1998 年 2 月制定的一种通用语言规范<sup>[1]</sup>, XML 是 SGML(标准通用标记语言)的一个子集, 非常适用于异构数据源间的数据交换, 而且这种交换不以预先规定一组数据结构定义为前提, 具有很强的开放性, 形式与内容分离、遵循严格的语法要求, 以及对多语种的支持, 为方便地跨平台跨

收稿日期: 2006-03-13

基金项目: 铁道部资助项目(Z2005-034)

作者简介: 吴静(1981-), 女, 江苏人, 硕士研究生, 主要从事计算机网络管理研究。

地域异构应用间的协同工作及数据搜索等提供了技术支持。XML 是一种元标记语言,所谓“元标记”就是开发者可以根据自己的需要定义自己的标记,任何满足 XML 命名规则的名称都可以标记。作为一种可扩展性标记语言,其自描述性使其非常适用于不同应用间的数据交换,而且这种交换是不以预先规定一组数据结构定义为前提。XML 也是一种语义/结构化语言。它描述了文档的结构和语义。

XML 包含 3 个要素:DTD 或者 XML Schema 规定了 XML 文件的逻辑结构,定义了 XML 文件的元素、元素的属性以及元素和元素属性的关系;XSL 用于规定 XML 文档呈现的式样;Xlink 进一步扩展目前 Web 上已有的简单链接。XML 提供描述不同类型数据的标准格式,比如动车段(所)管理信息系统中的维修工单、数据库记录等,而且能一致而正确地解码、管理和显示信息。

## 2 中介器集成技术简介

基于中介器的信息集成系统模型(下文称为中介器模型),如图 1<sup>[2]</sup>所示。

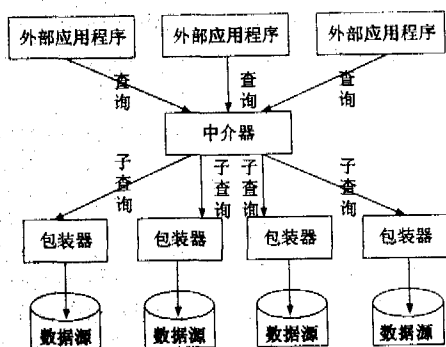


图 1 中介器信息集成模型

在中介器模型中<sup>[3]</sup>,并不将各数据源的数据集中存放,而是通过包装器(Wrapper)/中介器(Mediator)结构满足上层集成应用的需求。信息集成系统通过中介模式将各数据源的数据集成起来,而数据仍存储在局部数据源中,通过各数据源的包装器对数据进行转换使之符合中介模式。用户的查询基于中介模式,不必知道每个数据源的特点。中介器将来自外部应用程序的查询转换为基于各局部数据源模式的子查询交给包装器,包装器再将不同数据源的数据组织模式转换为信息集成系统的通用模式交给上层的中介器,最后由中介器将结果集成并返回给用户。

## 3 异构数据源集成框架设计

综合上述两种技术,动车段(所)信息系统异构数据源的集成框架可以由 4 个层次组成:应用层、接口层、中间件层、数据层。集成框架图<sup>[4,5]</sup>如图 2 所示。

各层功能如下:

1) 应用层:位于框架的最上层,与信息系统用户直接交互,提供各类应用服务支持。例如:各动车组调度信息

查询、维修决策支持等。

2) 接口层:位于中间件层和应用层之间,为应用层提供如 Web Services, XQuery, XPath, JDBC 等 XML 数据访问接口。

3) 中间件层:该层是框架核心层,位于数据层和应用接口层之间,向上为动车段(所)信息系统提供统一的全局数据模式,向下协调各异构数据源。中间件层包括元数据管理模块、查询分解模块、查询结果合并模块、XML 存储库以及不同数据源与 XML 之间的转换模块,其中查询分解模块与查询结果合并模块实现了中介器的功能。

4) 信息源层:各类数据的来源,例如调度运用计划是文本文件、各信息查询需要从数据库中取出数据,还有各种维修数据表格等。

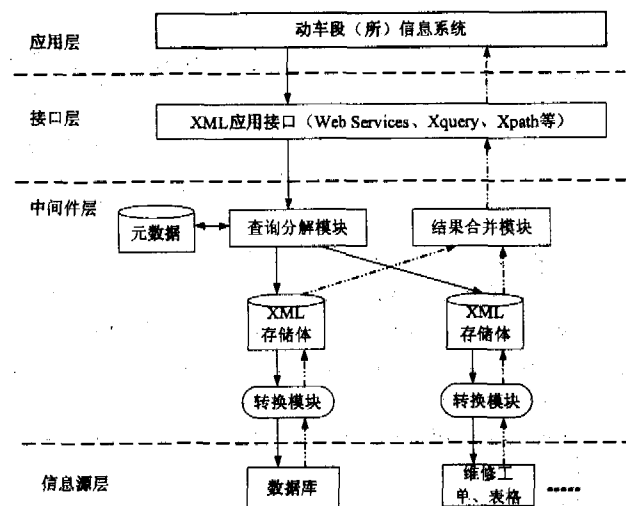


图 2 异构数据源集成框架图

以动车组运用信息查询为例,用户从应用层提出查询,要求显示 2006 年 2 月 28 日从北京段和武汉段分别开至南京段的所有动车组(假设北京段系统和武汉段系统使用不同数据库)。

查询请求被提出后,首先通过接口层中对应的接口程序将该请求以 XML 格式传递到中间件层。中间件层的查询分解模块根据收到的查询请求,在元数据管理模块中查找相应的映射规则,并据此对全局查询进行重写和分解,生成对各数据源的局部查询请求传递给各转换模块即包装器执行。比如该例中,查询被分解为针对北京段系统数据库的子查询和针对武汉段系统数据库的子查询。

接着由转换模块将各自管理的异构数据库的数据提取并转换成 XML 文档存储于本地 XML 存储库中,当收到针对本地数据库的局部查询请求后,即在 XML 存储库中执行查询请求,并上报查询结果。结果合并模块在收到返回的各数据库查询结果后,将查询结果进行整合,形成全局查询结果,通过应用接口层反馈给应用程序。

## 4 小 结

异构数据源集成与共享是很早就被提出的课题,但随

(下转第 152 页)

和正式)。这主要是因为,在大多数情况下,DTD 即可满足数据提取的需要,而无须支持更为复杂的 XML Schema。

⑥目前,模式映射方案在数据提取、异构系统之间进行数据交换的场合下具有较为广泛的应用。

## 2.5 模型映射方式

模型映射方式的过程大致如下:

(1)首先要定义一个类似于 DOM 的通用 XML 文档模型。一般而言,这个模型是一个树状结构。

(2)其次,需要设计一个(或者多个)能够存储这个模型的关系模式。

(3)最后,需要提供一个算法能够将具体的 XML 文档转换为该模型的实例。并将该实例分解存储到数据库相应的表中去。

该方式与文件和 CLOB/BLOB 方式有类似之处:在逻辑上,它们都将 XML 文档视为一个基本存取单位。但是模型映射中的 XML 文档是经过解析的高度结构化的数据,关系数据库可以把这些结构信息保存下来。文件和 CLOB/BLOB 方式中的 XML 文档则是未经过解析的文本块。这就决定了模型映射可以支持基于结构和类型的查询。

该方式操作过程如图 1 所示。

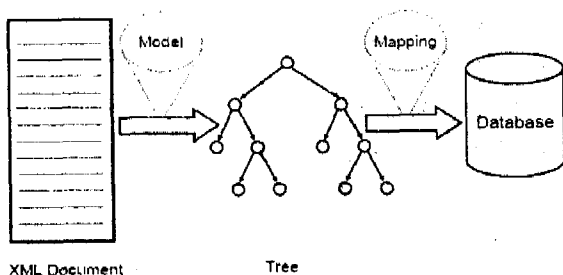


图 1 模型映射方式原理图

模型映射方式的特点:

- ①映射后,表的数目和字段都是确定的。
- ②可以保证数据的完整性,不会丢失数据。
- ③可以支持基于内容和结构的查询。

④对无模式的文档也适用。

另外,由于模型映射完全不涉及到模式;不能充分利用 XML 模式带来的优点;而且由于大多数模型抽象度比较高,存储模式相对呆板,在其上实现查询语言也有一定的难度。

## 3 结束语

XML 存储系统是 XML 数据库的关键。但是,由于 XML 数据本质上是一种自描述的半结构化数据,它的数据模型不同于以往的层次、关系、面向对象数据模型,已有的数据库技术和查询语言,不能完全适应于新的应用需求。目前,还没有一个能够得到广泛接受和认可的 XML 存储方案,所以有必要对该问题进行更深一步的研究。

## 参考文献:

- [1] 胡锡伟,陈仲委. Oracle 数据库的 XML 存储技术研究[J]. 计算机工程与设计,2005(5):179-181.
- [2] 徐德智,吴敏,赖同庆. XML 模式、查询和存储技术扫描[J]. 计算机工程与科学,2003(3):22-25.
- [3] ORACLE XML DB (An Oracle Technical White Paper [EB/OL]. 2004. <http://www.oracle.com/technology/tech/xml/xmlldb/Current/NewFeatures.pdf>.
- [4] IBM Informix Dynamic Server Getting Started Guide [EB/OL]. 2003. <http://www-306.ibm.com/software/data/informix/pubs/library/interim/ct1t1na.pdf>.
- [5] McHugh J, Abiteboul S, Goldman R, et al. Lore: A Database Management System for Semistructured Data [J]. SIGMOD Record, 1997, 26(3):54-66.
- [6] Jagadish H V, Shurug AL - Khalifa. TIMBER: A Native XML Database [R]. USA: University of Michigan, 2002.
- [7] Kanne C C, Moerkotte G. Efficient Storage of XML data [C]//In Proceedings of 16th ICDE. San Diego, California, USA: [s. n.], 2000.
- [8] 许卓明,刘琴,董逸生. 基于关系数据库的 XML 存储技术评述[J]. 计算机工程与应用,2003(21):197-201.

(上接第 149 页)

着应用环境的不断改变,这个课题也需要不断发展,如何采用合理的技术实现适应当前环境的数据集成将是一个重要的问题。

而近几年来来的发展趋势证明,XML 具有的简单性、规范性、开放性、灵活性和可扩充性等优点,能够有效地实现不同领域异构资源的集成,因此文中提出的 XML 技术在动车段(所)信息集成共享中的应用模式,能使动车段(所)内或动车段(所)间多种业务应用子系统、多种异构数据源并存,并实现异构数据源的动态及时、互访和信息的综合利用,能够满足组织低成本、阶段性、可扩展性动车段(所)信息系统建设的需要。

## 参考文献:

- [1] 王忠群,谢晓东. 基于 XML 的异构数据源的集成研究[J]. 安徽工程科技学院学报,2003,18(4):37-43.
- [2] 李伏欣. 铁路信息共享平台技术初探[J]. 中国铁道科学, 2002,23(5):29-35.
- [3] 孟小峰,周龙骧,王珊. 数据库技术发展趋势[J]. 软件学报,2004,15(12):1822-1836.
- [4] 李军怀,周明全,耿国华,等. XML 在异构数据集成中的应用研究[J]. 计算机应用,2002,22(9):10-12.
- [5] 赵毅,王浩然,庄冠华,等. 一种基于 XML 的数据集成系统框架及其应用[J]. 计算机工程与应用,2005(26):181-183.