

企业电子商务异构数据集成模型研究

沈智慧, 李隐璞

(湖南工业大学 计算机与通信学院, 湖南 株洲 412008)

摘要: 通过研究 XML 在数据集成方面的相关技术, 提出了一个基于 XML 企业电子商务异构数据集成框架, 可以无缝地与面向对象系统、数据仓库系统等进行接口, 能够较好地实现系统之间动态资源共享和交流, 为企业之间对各种分布异构数据库资源提供了全局、统一、动态和高效的访问。

关键词: XML; 电子商务; 异构数据; 数据集成

中图分类号: TP212.14

文献标识码: A

文章编号: 1673-9833(2008)04-0030-04

Research on the Model of Enterprise Electronic Commerce Isomerism Data Integration

Shen Zhihui, Li Yinpu

(School of Computer and Communication, Hunan University of Technology, Zhuzhou Hunan 412008, China)

Abstract: In view of related technology of XML studies in data integration, a frame based on the XML enterprise electronic commerce isomerism data integration is proposed. This frame can realize the dynamic resources sharing and exchanging between systems and also can provide the overall situation, unification and the highly effective visits for enterprises with every kind of distributed isomerism database resource by the seamless object-oriented system and the interface of the warehouse data systems.

Key words: XML; electronic commerce; isomerism data; data integration

0 引言

电子商务是随 Internet 及相关技术的发展而产生的一种现代商务方式, 是融计算机技术、通信技术于一体, 利用现代电子信息技术、特别是网络技术所开展的一系列商务贸易活动的总称^[1]。通过计算机与通信技术将产品的生产销售、配送结算和广大客户消费者紧密联系在一起, 快捷、低成本完成整个商品的销售过程。电子商务改变了传统的市场营销方式, 企业对外广告宣传、信息发布也不再采用传统媒体, 而是利用 Internet 的 Web 技术。

由于各企业之间使用的是不同类型的电子商务系统, 各自存储自己的数据, 企业之间越来越需要对各种分布异构数据库资源提供全局、统一、动态和高效的访问。因为只有把这些数据信息联合汇总起来才能

发挥这些数据的最大效用, 所以需要利用数据集成技术将异构数据转化为可扩展标记语言 XML (Extensible Markup Language) 数据。因此, 信息资源的动态变化成为了数据集成研究的一个主题。

传统数据集成技术处理高度动态异构数据资源的能力还有很多缺陷, 无法满足企业电子商务数据交换的要求。本文通过研究 XML 在数据集成方面的相关技术, 提出了一个基于 XML 企业电子商务异构数据集成模型, 以解决 Web 中各种信息系统之间的接口问题。通过接口不但可以抽取网络上有意义的 XML 信息, 在关系数据库中进行有效的管理, 而且还提供了相应的策略可以无缝地与面向对象系统、数据仓库系统等进行接口, 能够较好地实现系统之间动态资源共享和交流, 具有良好的可扩充性。

收稿日期: 2008-04-18

作者简介: 沈智慧 (1952-), 女, 湖南湘潭人, 湖南工业大学教授, 主要研究方向为计算机应用, 计算机数据库和网络技术。

1 基本概念

XML是W3C(World Wide Web Consortium)组织提出的一个Internet上数据表示与数据交换的新标准,是一种灵活的、动态的、可扩展的、并得到广泛支持的数据表示语言^[2]。XML结合了SGML和HTML的优点并消除了其中的缺点。由于其特有的自描述能力及结构化的特征,XML在Web数据交换、信息管理以及电子商务等领域具有重要的作用。

将信息按照文档模式来存放和管理是XML的一大特点,这种模式非常接近人类对信息的组织和管理模式。一方面XML提供了丰富的、自描述的、统一的信息表示,另一方面这种信息表示能够同时被人类和计算机所理解。信息一旦采用XML来表示,那它就可以方便地被其它系统和应用共享和利用。XML的完美之处在于它能表示任何类型的信息,就像TCP/IP实现全世界范围内不同网络间的信息通讯,XML使得全世界的信息变得开放和通用。XML丰富灵活的表现形式为异构数据的集成与共享提供了重要的基础。

目前,关于异构数据源信息集成的问题已有很多研究,如Haihong Dai提出了面向对象的框架,基于属性等价理论;Florescu则提出利用概率统计信息集成多个数据源^[3]。另外,虽然XML Schema中已经提供了定义XML数据完整性约束的机制,但是在数据交换过程中,原来关系数据库中的各种语义约束,不能通过交换过程转换为XML上的语义约束。本文利用XML作为主要方法,利用W3C的XSLT(Extensible Stylesheet Language Translation)定义不同元素之间的影射关系,实现多数据源的模式集成。

2 异构数据集成模型

数据集成的做法是采用某种合适的语言对数据源中的数据建模,并能够统一地表示查询和数据源的内容,最后通过合适的机制访问数据源(或实例化视图)来回答用户的查询^[4]。

本文研究的企业电子商务异构数据集成模型是基于XML的一个重要应用,其目的是为用户提供一个集成、一致的、能够对存储在多个异构的信息源中的数据进行访问的视图。系统能够将用户的查询转换成对相应的数据源访问,然后返回查询结果给用户,免除了目前从Web上手工收集异构信息数据的繁琐工作。

企业电子商务异构数据集成模型不但考虑了模式的集成,而且考虑了实际的数据集成。在这种允许更新操作的集成模型中,输入部分是源数据的集合(含概念模式和实际数据),而各个数据源可以是异构的和自治的。由于更新操作会改变数据源中的数据,因此,模型既要考虑如何消除异构性,也要求数据源和集成模

型之间以及不同数据源之间有更高的协作。

企业电子商务异构数据集成模型是在充分吸收国内外比较成熟的电子商务标准各方面的优点结合信息技术研究设计的,其体系结构包括商业伙伴协议规范、商业伙伴描述规范、模型流程规范、模型文档规范、模型公共元素库、模型注册查询规范和模型消息服务规范。模型如图1所示。

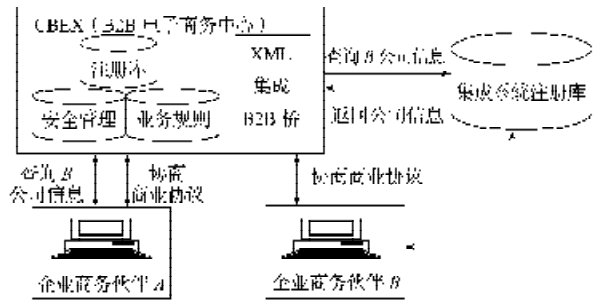


图1 企业电子商务异构数据集成模型

Fig.1 Enterprise electronic commerce isomerism data integration model

3 系统设计

系统总体流程如图2所示。

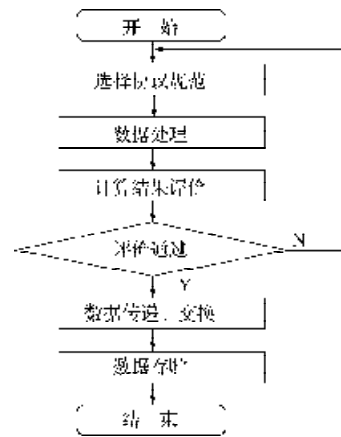


图2 系统总体流程图

Fig.2 System overall flow chart

早期的电子数据交换EDI(Electronic Data Interchange)和企业应用集成EAI(Enterprise Application Integration)等技术都已无法适应企业电子商务数据交换的动态需求,XML标准的提出和发展,为企业电子商务多样化数据提供了一种描述语言。由于XML几乎可以描述任何数据类型,从简单的整型到复杂的对象类型,所以,XML是一种较理想的数据集成的工具和异构数据源集成的接口语言。

要实现企业电子商务异构数据集成,首先必须对建立的各种规范进行选择,使用相应的算法实现数据处理,然后进行计算结果的评价和判断。如果计算结果评价获得通过,就可以进行数据的传递、交换及存

储, 实现对间各种分布异构数据库资源提供全局、统一、动态和高效的访问。

3.1 数据处理流程设计

基于 XML 的异构信息集成系统文档处理流程如图 3 所示。

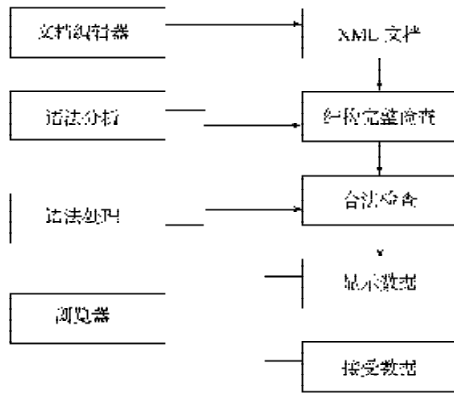


图 3 文档处理流程图

Fig. 3 Document processing flow chart

第一步, 生成 XML 文档。当前, XML 文档的表现方法有层叠样式表 (CSS) 和 XML 样式表语言 (XSL) 两种方法。采用层叠样式表 (CSS) 以类似文档的方式来表现 XML 文档片段, 需要先生成本格式的样式表, 然后再将样式表与 XML 文档相结合。

要实现用 XSL 表现 XML, 需要先生成 XSL, 然后再将它与 XML 文档相结合。XSL 分为 XSLT 和 XSL-FO (XSL-Formatting Object) 两个部分, XSLT 负责将 XML 的源代码转换为 XSLT, 而 XSL-FO 提供大量的格式化命令, 用来配合打印或屏幕显示。XSLT 的转换语法除了将 XSL-FO 服务转换成 FO 命令外, 还可以实现 XML 与 HTML、XML 与 XML、XML 与 XSL 语法之间的转换。这样, XSL 能使 Web 浏览器不需要再次与服务器进行交互通信, 就可以根据用户的需求改变文档的显示方式。

第二步, 进行 XML 文档解析。XML 的语法分析程序读取文档并检查其中包含的 XML 是否结构完整。如果文档通过了测试, 则处理程序就将文档转换为元素的树状结构。

解析有 2 种标准, 即 DOM 和 SAX: 在需要详细了解文档的结构、移动文档的组成部分或者多次使用文档信息等情况下, 使用 DOM 解析; 当需要从文档中抽取一些元素, 或者内存容量小等情况, 则使用 SAX 标准解析。

第三步, 文档的显示和用户交互。通过浏览器进行 XML 文档显示或转换成其它结构化数据。

3.2 数据的传递设计

将数据从源数据库转移到目的数据库的目的, 就是要在目的数据库中再现源数据库中信息的内容和结

构。数据库中的对象和 XML 文档都具有树状结构, 数据库中的对象树是由对象、子对象、属性和对象与后两者间的连结构成的。而 XML 文档的结构树则由元素、子元素 (subelement)、属性、PCDATA 和元素与后三者间的连结构成^[5]。本文就是采用 XML 文档种的对象树数据模型在 XML 文档和数据库之间传递数据。

1) XML 文档具有树状结构, 按深度有限的原则从树叶往树根搜索;

2) 每碰到 1 个内部节点 (叶节点和根节点除外), 查看数据库中对应的抽象数据类型是否存在, 若不存在, 则创建对应的抽象数据类型, 该抽象数据类型应将这个节点的所有直接子节点作为它的域;

3) 重复步骤 2), 直到根节点的直接子节点;

4) 创建基于次根节点所对应的抽象数据类型的对象, 该对象就是根节点对应的对象;

5) 将 XML 文档中的各元素似的属性值填入对象中的相应域。

举例, 企业电子商务异构数据集成系统接收到的销售单据文档可以被看成是 1 个对象树, 其中包括 5 个类: 即 Orders、SalesOrder、Customer、Line、Part。销售单据文档对象树如图 4 所示。

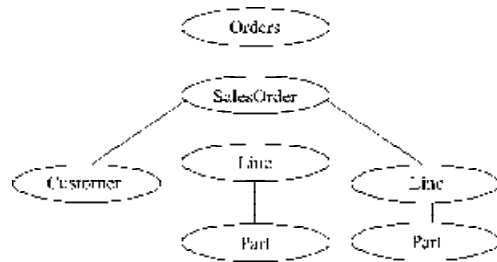


图 4 销售单据文档对象树

Fig. 4 Sales Order Document Object tree

当 1 个 XML 文档模型化处理成 1 个对象树时, 结点代表了元素或者元素的属性, 结点的祖孙、父子关系表达了元素之间的层次结构。如果 1 个元素只包含有 PCDATA, 如销售单据文档中的 CustName 元素, 它就能够被看作 1 个属性进行处理 (仅仅只包含有单独的数值)。销售单据文档中对 Description 元素的处理时, 尽管它在 XHTML 的 Form 中有混合的内容, 但是将 Description 元素看作 1 个单独的属性来处理会更有用些。

3.3 数据交换设计

数据交换的核心问题是要解决信息的标准化问题, 这不但可以提高人和机器对信息的理解识别, 而且能根据数据进行自动处理。由于数据交换在企业电子商务中具有核心作用, XML 的出现为企业电子商务带来了新的机遇与活力, 为信息的标准化提供了有力的工具。

与 HTML 相比, XML 在数据交互方面具有非常明

显的优势。第一, 它支持客户端在 2 个或多个异种数据库间进行操作; 第二, 它支持数据集视图, 即相同伴数据的多种表现形式; 第三, 它支持信息发现与智能检索。

基于 XML 企业电子商务异构数据集成模型数据交换过程描述如下: 在服务器端通过 Socket 建立监听组件, 当接收到客户端的连接请求后, 与客户端建立连接。然后, 客户端可以把查询请求通过 Socket 连接发送到服务器端。服务器端执行 ADO 执行数据库的查询操作, 并将查询的结果封装成 XML 的形式, 通过 Socket 连接发送到客户端。客户端则通过解析收到 XML 字符串, 得到所需的信息。

3.4 数据存储设计

XML 在数据应用方面具有易表义、跨平台等优势, 但也需要一种有效的存贮、检索和修改等方面的管理机制。在数据管理方面, 数据库尤其是关系数据库具有与其它方式(包括普通文件、对象数据库)无法比拟的优越性。在数据应用方面, 尤其在 Web 信息共享及异构数据交换方面, XML 又具有与其它技术无法比拟的优势。

企业电子商务系统都是和数据库相关联的。当把 XML 数据存储到数据库时, 为了在数据库和 XML 文档之间传递数据, 就必须在二者结构之间建立映射。本文采用在 XML 数据存储模块中嵌入命令, 并用数据传输中间件来进行处理:

```
<? xml version= "1.0" ?>
< FlightInfo >
< Intro > The following flights have available seats:
< /Intro >
< SelectStmt > SELECT Airline, FltNumber, Depart ,
Arrive FROM Flights < /SelectStmt >
< Conclude > We hope one of these meets your needs
< /Conclude >
< /FlightInfo >
```

本文在数据传输中间件嵌入了一个 SELECT 语句。当用数据传输中间件进行处理的时候, 每一个 SELECT 语句都会被它的结果所替代。用 XML 格式化形式表现为:

```
<? xml version= "1.0" ?>
< FlightInfo >
< Intro > The following flights have available seats:
< /Intro >
```

```
< Flights >
< Row >
< Airline > ACME < /Airline >
< FltNumber > 358 < /FltNumber >
< Depart > Dec 12, 2006 13: 43 < /Depart >
< Arrive > Dec 13, 2006 01: 21 < /Arrive >
< /Row >
...
< /Flights >
< Conclude > We hope one of these meets your needs
< /Conclude >
< /FlightInfo >
```

4 结语

随着全球化经济竞争的加剧、信息化技术的飞速发展, 对企业电子商务应用系统开发技术提出了新的要求。由于早期的电子数据交换 EDI 和企业应用集成 EAI 等技术已经无法适应企业电子商务系统的动态需求, 而构建基于 XML 技术的企业电子商务异构信息集成系统正是解决这一问题的最佳手段。

本文描述的基于 XML 企业电子商务异构信息集成模型, 充分研究和利用了 XML 及其相关技术, 能够较好地实现动态资源共享, 解决了数据库变更所带来的服务自适应性问题, 能更好地支持企业间异构数据的集成, 不但可以直接构建于已有的标准协议之上, 而且能使数据交换更加简单、廉价和方便。这必将推动网络数据交流、数据集成、数据共享等数据管理技术的不断完善和快速发展。

参考文献

- [1] 邵兵家. 电子商务概论[M]. 北京: 高等教育出版社, 2006.
- [2] 徐宝文, 陈 凌, 周晓宇. 程序设计语言与教学进展[M]. 北京: 电子工业出版社, 2003.
- [3] 王晓玲, 董逸生. 面向Web的异构信息系统集成方案[J]. 计算机科学, 2003, 10(30): 267.
- [4] 谢丽聪, 白清源, 余建家. 数据集成系统的三层体系结构及查询改写算法的改进[J]. 计算机科学, 2003, 10(30): 255.
- [5] 赵怀璧. XML 文档到关系数据库的转换实现[M]. 北京: 电子工业出版社, 2003.

(责任编辑: 罗立宇)