

基于模型驱动的 XML 与数据库双向映射技术

孙宏伟 张树生 周竞涛 王 静

(西北工业大学 CAD/CAM 国家专业实验室,西安 710072)

E-mail shw2000@263.net

摘 要 XML(eXtensible Markup Language)不但是第二代 Web 的核心,同时也是异构应用数据交换的中性载体,基于 XML 的各种应用研究已经广泛展开。文章在比较 XML 和数据库各自特点的基础上,提出 XML 和数据库相结合实现异构应用数据交换问题,研究基于模型驱动的 XML 与数据库的双向映射方法,讨论双向映射过程中存在的问题,并结合典型例子,对该方法进行验证。为实现 XML 和数据库之间的数据有效传输提供所需的关键支撑技术。

关键词 XML 模型驱动 数据库 双向映像 数据交换

文章编号 1002-8331-(2002)04-0025-03 文献标识码 A 中图分类号 TP391

The Technology of Bidirectional Mapping Between XML and Database Based on Model-drive

Sun Hongwei Zhang Shusheng Zhou Jingtao Wang Jing

(National Specialty Laboratory of CAD/CAM, Northwestern Polytechnical University, Xi'an 710072)

Abstract : XML(eXtensible Markup Language) is not only the core technology of the second Web but also the neutral carrier of data exchange among isomorous application and diverse research works based on XML application have been widely taken out. This paper first compares XML with databases and puts forward the opinion of combining XML with database to realize the exchange among isomorous applications. Then it also studies the bidirectional mapping technique between XML and databases based on model-drive and discusses the possible problems. With the typical example this technique is validated and it will lend deep-seated technical support to the data transformation between XML and databases.

Keywords : eXtensible Markup Language(XML), Model-Drive, Database, Bidirectional Mapping, Data Exchanging

1 引言

XML 作为一种可扩展标记语言,其自描述性使得 XML 本身非常适用于异构应用间的数据交换,而且这种交换不以预先规定的数据结构定义为前提,因此具备很强的开放性和广阔的应用前景。而现有的几乎所有大型应用系统都是和数据库相关联的,所以以 XML 中性载体进行数据交换和信息共享必然要和数据库相联系,同时,基于 XML 的数据交换,实现数据库的 XML 数据双向存取,才能将 XML 数据同具体应用程序集成,进而使之同现有的业务规则相结合,最后真正实现基于 XML 的分布式数据交换与信息共享。所以,XML 技术和数据库技术的结合就成为必然的趋势,在 XML 文档和数据库之间实现数据的双向传输就成为 XML 技术应用研究的热点之一,基于 XML 的数据库应用也迅速发展起来,具体表现有^[1]:

(1)应用中间件技术(Middleware),例如 Microsoft 的中间件产品 ADO、IBM 的 DatabaseDom;

(2)与 XML 相结合能驱动 XML 的数据库(XML-Enabled Databases),例如 Microsoft 的 Microsoft SQL Server 2000、IBM 的 DB2 XML Extender and DB2 Text Extender;

(3)原始的 XML 数据库(Native XML Databases),例如 Birdstep 公司的 Birdstep DataBase Engine;

(4)XML 服务器(XML Servers),例如 Wizen Software 的 PowerXML Server and PowerXML Pro;

(5)XML 应用服务器(XML Application Servers),例如 InterSystems Corp 的 Cache;

(6)内容管理系统(Content Management Systems);

(7)可持久化的 DOM 实现(Persistent DOM Implementations);

(8)XML 询问机(XML Query Engines)等。

2 XML 与数据库

XML 应用内容、结构、表现三者的分离技术,实现了文档结构、文档内容、及其表现形式的分离。较传统意义上的文档具有很大的灵活性,使得 XML 文档逐步成为异构应用间系统集成和数据交换的中性载体。

XML 本身并不是数据库,从严格意义上来说,XML 仅仅是指 XML 文档。因为尽管一个 XML 文档包含数据,但是如果通过其它的软件来进行诸如文档解析、数据合成等数据处理,它本身只不过是一个文本文件。所以,相对于数据库系统的意义,XML 文档和其他文本文件并没有什么区别。

如果从更为宽泛的意义来讲,当 XML 是指 XML 文档以及

基金项目:国家 863 高技术研究发展计划基金资助(编号:863-551-942-07)

作者简介:孙宏伟,男,西北工业大学 CAD/CAM 国家专业实验室博士研究生,研究方向:数据库技术、企业信息系统、PDM。张树生,博士、教授、博士生导师,主要研究方向:网络数据库、PDM、神经网络。周竞涛、王静均为西北工业大学 CAD/CAM 国家专业实验室博士研究生。

所有相关的 XML 的工具和技术时,可以把整个 XML 看成是一个数据库系统,之所以这样认为,是因为 XML 提供了许多数据库中所必要的数据处理功能。XML 文档本身可以看成是数据库中的数据区,DTD 或者 Schemas 可以看成是数据库模式设计,包括关系数据库的关系模式(Relational Schemas)和面向对象数据库(Object Schemas);XQL 或 XML-QL 可以看成是数据库查询语言 SQL;XML 的解析方法 SAX(Simple API for XML)或 DOM(Document Object Model)可以看成是数据库处理工具。然而,相对传统的管理数据的数据库技术,XML 还是缺少数据库所必须的一些东西,比如有效的存储过程、索引结构、安全性、事务处理、数据完整性、触发器、多用户处理机制等等。如果在数据量一般、用户较少、性能要求不高的环境下可以把 XML 当作数据库来使用;而在大多产品的环境中,要求有许多的用户使用、需要严格的数据完整性并且对性能有很高的要求,XML 就不能胜任了。而且,考虑到象 dBase 和 Access 等数据库既便宜又十分易用,因此 XML 很少有理由充当数据库的角色。所以 XML 不能也不会代替数据库技术。

但是 XML 本身较数据库又具有很多优点:

(1)开放性,一般数据库系统数据库结构的不公开导致了应用间数据交换的瓶颈,而通过 DOM 或 SAX 的 XML 解析技术可以分析 XML 的 Schemas 或 DTD 获得 XML 的文档组织结构,实现异构应用间数据交换;

(2)移动性,一般大型的数据库系统的移动非常费时费事,而 XML 的移动非常容易,只需移动文件而已。

(3)适应性,XML 可处理 Unicode 字符集,大多数数据库系统仅支持 ASCII 码。

因此 XML 和数据库各有优势,XML 作为异构应用间数据交换的中性载体必需和应用内部数据库进行数据交换。

3 基于模型驱动的双向映射技术

为了在数据库和 XML 文档之间传递数据,必须在文档结构和数据库结构之间建立映射,这种映射可以分为模板驱动和模型驱动。基于模板驱动的映射需要在一个模板中嵌入带参数的 SQL 命令,并用数据传输诸如中间件等实体软件进行处理,基于模板驱动的映射是一种浅层映射,以 XML 内嵌的 SQL 执行的数据结果集为依据,不涉及数据库赖以存在的关系模式或对象模式,值得注意的是,当前诸如 Microsoft SQL Server 2000 等大多数的产品都属于模板映射,只能应用于在关系数据库和 XML 文档之间传递数据,并不支持面向对象数据库或对象数据库和 XML 之间的数据传输。

基于模型驱动的映射,当把数据从数据库传送到 XML 文档或把数据从 XML 文档传送到数据库时,是用一个具体的模型实现的,而不是仅仅依赖内嵌 SQL 命令,关系数据库的理论依据是关系模型,面向对象数据库的依据是对象模型,而 XML 文档的依据是 XML Schemas 或 DTD。基于模型驱动实现数据在数据库和 XML 文档间的双向传输关键是在数据库模式和 DTD 或 XML Schemas 之间的建立双向映射。具体的思路是把一个 XML 文档表示为由数据对象构成的树,每一个元素类型和对象模式中的对象相对应,主要在面向对象数据库中使用,通过传统的关系-对象模型也可以映射到关系模式中。

模型驱动的映射是一种深层映射,是模式和模式之间的映射。实际上在 XML 和数据库之间进行数据转化的时候,需要考

虑两个逆反的过程:一个是从数据库模式中产生 DTD,另外一个是根据 DTD 生成数据库模式。

3.1 DTD 到数据库模式的映射

按照上述思路:从一个 DTD 中生成一个数据库模式(Database Schemas)的步骤如下:

(1)一个 DTD 生成一个关系模式,在此基础上建立关系数据库。

(2)对 DTD 中的每一个元素,产生关系数据库的一个表和一个主键列。

(3)对每一个有混合内容的元素,产生一个独立的表,用来存储 PCDATA(PCDATA 表示已解析字符数据,在这里指不需再解析的最小数据单元),并通过父表的主键和父表相联。

(4)对元素类型中的每一个单一值的属性,即对具有只有 PCDATA 内容的按顺序出现的子元素,产生一个单独的列,如果子元素类型或者值是可以选择,该列就应该可以允许为 NULL 类型。

(5)对有多值的属性和可以出现多次的 PCDATA 类子元素,需要创建一个单独的表来存储这些值,并通过父表的主键和父表相连。

(6)对每一个包含元素或者混合内容的子元素来说,通过父表的主键把父元素和子元素连接起来。

3.2 数据库模式到 DTD 的映射

从一个 DTD 中生成一个数据库模式的步骤如下:

(1)对每一个表,创建一个元素。

(2)对表中的每一列,创建一个属性或者是一个只有 PCDATA 内容的子元素。

(3)根据表中的每一主键/外键关系,创建该表元素的子元素。

XML 缺少诸如存储过程、索引、事务处理、触发器等一些机制,所以这种映射属于不等价映射。

3.3 双向映射过程中的若干问题讨论

由于 XML 本身是一种文档,数据库有着严密的组织结构,在映射中存在很多问题,这里列举一些,给以说明。

(1)数据类型:XML 不支持数据类型,也就是说,在 XML 文档中的所有数据都是文本,即使数据本身代表了另外一种数据类型,比如日期或者整数,可以加入数据类型标记来标识,通过数据传输中间件把数据转化为对应的数据类型。

(2)二进制数据处理:有两种通常的方法存储 XML 文档中的二进制数据,未经过解析的实体(unparsed entities)。

(3)Null 类型处理:在数据库中, NULL 表示该数据“不知道”,它和 0 或空字符串是不一样的,XML 也支持 NULL 的概念。当映射一个 XML 文档的结构到数据库或者根据数据库内容生成 XML 文档的时候,需要考虑可选元素类型和属性跟可 NULL 列之间的映射。

(4)字符集:XML 文档可以包含任何 Unicode 字符,大多数数据库并不支持 Unicode。因此数据包括非 ASCII 字符的时候,需要注意中间件对这些字符的处理。

4 映射实例

根据上述规则,完成了一个用户订单实例的映射,数据库模式如图 1 所示,对应的 DTD 如下:

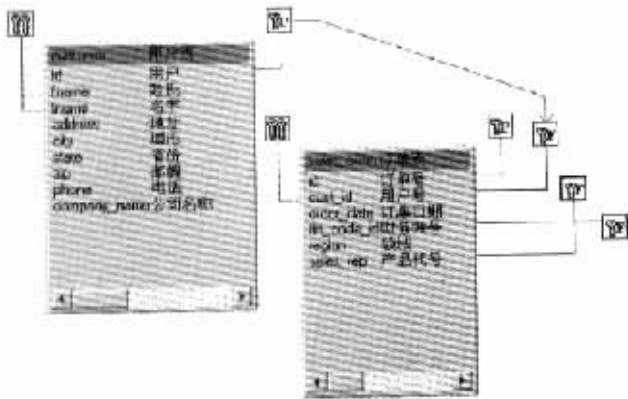


图 1 用户-订单数据库模式

```
<!--Generated 2001-6-6 18 30 45 by EditML Pro ,Version 2.0-->
<!ELEMENT sales_order( Record* )>
<!ELEMENT Record
(id cust_id order_date fin_code_id region sales_rep Foreignkey )>
<!ELEMENT id( #PCDATA )>
<!ELEMENT cust_id( #PCDATA )>
<!ELEMENT order_date( #PCDATA )>
<!ELEMENT fin_code_id( #PCDATA )>
<!ELEMENT region( #PCDATA )>
<!ELEMENT sales_rep( #PCDATA )>
<!ELEMENT Foreignkey_group( Foreignkey )*>
```

(上接 16 页)

4.4 扩展过程性能

表 4 扩展过程性能

	8KB	1024KB
平凡方法	3 小时 52 分 49 秒	1 小时 30 分 49 秒
散布法	3 分 7 秒	8 分 55 秒

这里测试了使用两种扩展方法扩展 2 磁盘阵列到 3 磁盘阵列所需时间,实验使用的是 IBM DDYS-T36950,容量 36GB,实验平台为 RedHat Linux 7.1。表 4 给出了结果,显见,散布法由于数据移动量小,磁头顺序移动,表现出了高出平凡方法一两个数量级的性能优势。

5 总结

随着计算机技术和应用需求的快速发展,在实际应用中人们常常遇到系统升级扩展的情况,对存储系统的扩展就涉及到 RAID 结构扩展问题。对于这一问题,目前相关的研究非常少。在一些工业产品中提供了 RAID 结构扩展功能,但基本都是采用最简单的平凡方法。这种方法的最大缺点是扩展过程性能太差,对系统正常工作有较大影响。针对平凡方法的这一缺点,作者借鉴“热闲置空间散布法”,提出了“新增空间散布”扩展方法。这种方法减少了数据移动,因此在扩展性能上比平凡方法有较大提高。虽然破坏了标准 RAID0 结构,但通过分析和仿真实验表明,在常见的大多数应用所采用的多用户并发访问模式下,其性能是完全可以与 RAID0 结构相比的。只是对于某些特定的应用,新结构的性能才有较大下降。因此,散布法应该是一种具有一定应用价值的快速 RAID 扩展方法。

该文仿真实验采用的是人工生成请求的模式,更好的方式

```
<!ATTLIST Foreignkey( #PCDATA )>
<!ATTLIST Foreignkey Table_name ID #REQUIRED>
<!ATTLIST Foreignkey id( #PCDATA )>
<!ATTLIST Foreignkey fname( #PCDATA )>
<!ATTLIST Foreignkey lname( #PCDATA )>
<!ATTLIST Foreignkey address( #PCDATA )>
<!ATTLIST Foreignkey city( #PCDATA )>
<!ATTLIST Foreignkey state( #PCDATA )>
<!ATTLIST Foreignkey zip( #PCDATA )>
<!ATTLIST Foreignkey phone( #PCDATA )>
<!ATTLIST Foreignkey company_name( #PCDATA )>
```

5 结束语

以 XML 为中性载体,基于 XML 和数据库的有机结合进行数据传输,是解决异构应用间信息共享的一条有效途径。文章提出的基于模型驱动的 XML 和数据库双向映射,为实现 XML 和数据库之间的数据有效传输提供了所需的关键支撑技术。(收稿日期 2001 年 7 月)

参考文献

- 1.Database Middleware.http://www.javaworld.com/javaworld/tools/jw-tools-datamid.html
- 2.XML and Database.http://www.rpbouret.com/xml/
- 3.XML Database Product.http://www.rpbouret.com/xml/
- 4.Mapping W3C Schemas to Object Schemas to Relational Schemas.http://www.rpbouret.com/xml/XMLAndDatabases.htm

可能是使用跟踪实际系统得到的跟踪日志,这样能更好地测试 RAID 结构的实际性能,这是作者今后的工作之一。散布法在某些情况下性能确实较差,因此探索更好的 RAID 结构也是今后工作的重要内容。今后的另一项工作是探索其它 RAID 结构的扩展方法,不同 RAID 结构差异较大,散布法是否适用于其它 RAID 结构,是否需要改进,是否需要新的方法,都要进行研究。另外,在系统升级扩展中,常遇到设备型号不一致的情况,因此异类磁盘阵列的扩展问题是今后研究的又一课题。

(收稿日期 2001 年 12 月)

参考文献

- 1.David A Patterson et al.A Case for Redundant Arrays of Inexpensive Disks[RAID][C].Proceedings of the 1988 ACM SIGMOD International Conference on Management of Data,1988
- 2.John Chandy A L Narasimha Reddy.Failure Evaluation of Disk Array Organizations[C].Proceedings of the 13th International Conference on Distributed Computing Systems,1993
- 3.Ethan L Miller Randy H Katz.RAMA: An Easy-to-Use, High-Performance Parallel File System[J].Parallel Computing,1997,23(4-5)
- 4.W V Courtright II et al.RAIDframe: Rapid Prototyping for Disk Arrays[C].Proceedings of the 1996 ACM SIGMETRICS Conference on Measurement and Modeling of Computer Systems,1996
- 5.T Cortes J Labarta.A Case For Heterogeneous Disk Arrays[C].1st IEEE International Conference on Cluster Computing 2000
- 6.M Holland et al.Architectures and Algorithms for On-Line Failure Recovery In Redundant Disk Arrays[J].Journal of Distributed and Parallel Databases,1994,2(3)

作者: 孙宏伟, 张树生, 周竞涛, 王静
作者单位: 西北工业大学CAD/CAM国家专业实验室, 西安, 710072
刊名: 计算机工程与应用 **ISTIC PKU**
英文刊名: COMPUTER ENGINEERING AND APPLICATIONS
年, 卷(期): 2002, 38 (4)
被引用次数: 43次

参考文献(4条)

1. [Database Middleware](#)
2. [XML and Database](#)
3. [XML Database Product](#)
4. [Mapping W3C Schemas to Object Schemas to Relational Schemas](#)

本文读者也读过(10条)

1. 徐小平, 董威, 齐治昌 [UML2.0的发展历程及特征](#)[会议论文]-2003
2. 方翔, 袁国栋, 李伟生 [从XML模式到关系模式的影射](#)[期刊论文]-[计算机工程与应用](#)2002, 38 (19)
3. 朱虹, 罗荣, ZHU Hong, LUO Rong [XML数据库强制访问控制策略研究](#)[期刊论文]-[计算机工程与科学](#)2007, 29 (9)
4. 龚进峰, 逢红 [几种新兴汽车电子技术](#)[期刊论文]-[世界汽车](#)2004 (2)
5. 陈敏, 姜宏 [基于MDA的工作流模型的研究](#)[会议论文]-2006
6. 廖永为, 杨丹, 朱勇 [基于UML/XML的ASIP系统级设计研究](#)[期刊论文]-[大众科技](#)2010 (3)
7. 王金波 [应用模型驱动开发方法的1553B总线仿真终端的开发](#)[学位论文]2006
8. 杜军威 [面向Web的异构数据库交换平台的设计与实现](#)[学位论文]2005
9. 宋文鹏, 侯敏, 何瑞, 贺红卫 [基于状态机的模型驱动软件开发技术研究](#)[会议论文]-2010
10. 崔萌, 袁海, 史耀馨, 李宣东, 郑国梁 [一种基于MDA的UML顺序图到状态图的转换方法](#)[期刊论文]-[南京大学学报\(自然科学版\)](#)2004, 40 (4)

引证文献(43条)

1. 贺晓萍 [基于XML的数据库存取技术](#)[期刊论文]-[湖南理工学院学报\(自然科学版\)](#) 2007 (1)
2. 王伟达, 芦东昕, 孟照星 [关系数据库与XML的双向数据传输的机制与实现](#)[期刊论文]-[计算机应用研究](#) 2005 (2)
3. 杨慧慧 [XML异构数据库间数据安全交换技术研究](#)[期刊论文]-[泰山医学院学报](#) 2011 (2)
4. 余文芳, 荆泽泉 [基于XML和WebServices的异构数据集成研究](#)[期刊论文]-[科技信息](#) 2011 (27)
5. 关辉 [异构数据库间数据安全交换技术研究](#)[期刊论文]-[电脑知识与技术](#) 2009 (1)
6. 潘雪峰 [基于XML的生物信息数据整合模型](#)[期刊论文]-[现代电子技术](#) 2009 (20)
7. 宋长新 [基于XML的生物信息平台构架](#)[期刊论文]-[青海师范大学学报\(自然科学版\)](#) 2008 (3)
8. 刘卓, 张晶泊 [基于XML的一种多Agent异构数据查询模型设计](#)[期刊论文]-[大连海事大学学报](#) 2007 (z1)
9. 姜莉 [基于XML技术的多Agent模型的设计与实现](#)[期刊论文]-[科技经济市场](#) 2007 (5)
10. 李颖富, 杨宏, 苑津莎 [基于Web服务/XML的配电网信息发布集成研究](#)[期刊论文]-[电力系统通信](#) 2006 (3)
11. 熊辉 [异构数据库连接平台的分析与设计](#)[期刊论文]-[科技信息](#) 2011 (21)
12. 熊辉 [异构数据库连接平台中的数据映射](#)[期刊论文]-[科技信息](#) 2011 (3)
13. 吕红, 赵媛 [企业信息网格中XUL-SQL技术的研究](#)[期刊论文]-[微计算机应用](#) 2010 (6)
14. 赵媛, 吕红, 任颖, 吕海燕 [基于XML View的远程信息修改研究](#)[期刊论文]-[自动化技术与应用](#) 2009 (8)
15. 宋长新 [基于XML的生物信息数据模型](#)[期刊论文]-[青海师范大学学报\(自然科学版\)](#) 2007 (4)

16. 林万辉, 童水光 [基于XML的PDM与AutoCAD属性双向映射](#)[期刊论文]-[机械制造](#) 2007(8)
17. 刘扬, 赵文吉, 宫辉力 [关于异构数据连接的研究](#)[期刊论文]-[首都师范大学学报\(自然科学版\)](#) 2005(1)
18. 李冠宇, 黄海, 张俊 [XML在异构数据集成方面的应用](#)[期刊论文]-[交通与计算机](#) 2002(5)
19. 朱彬, 徐俊刚 [基于SOA架构的异构数据库整合框架](#)[期刊论文]-[电子技术](#) 2011(6)
20. 胡鑫 [MES与ERP系统集成接口的研究与设计](#)[学位论文]硕士 2006
21. 熊海军, 王保义, 张少敏, 张琦 [基于XML的Web数据集成研究](#)[期刊论文]-[华北电力大学学报](#) 2004(5)
22. 林莉, 傅晓娟, 陈丽丽 [XML数据交换技术在高校信息化中的应用](#)[期刊论文]-[电脑知识与技术](#) 2012(3)
23. 赵媛, 吕红, 杨阿桃, 王丽娜, 张杰 [分布式信息集成中XUL-SQL问题的研究](#)[期刊论文]-[微电子学与计算机](#) 2011(8)
24. 张飞, 张建 [数据库和内容仓库异构数据共享解决方案](#)[期刊论文]-[计算机工程与设计](#) 2009(9)
25. 侯占伟, 莫林, 郑华, 翟海霞 [基于SOA的数据库中间件的研究与设计](#)[期刊论文]-[计算机应用研究](#) 2007(6)
26. 关辉 [异构数据库间数据交换技术研究与实现](#)[期刊论文]-[电脑知识与技术\(学术交流\)](#) 2007(19)
27. 王娟 [基于XML的异构信息处理系统及应用](#)[学位论文]硕士 2005
28. 张建, 刘更, 贺朝霞, 王海伟 [基于XML实现Java内容仓库和关系型数据库的双向映射](#)[期刊论文]-[计算机应用研究](#) 2009(1)
29. 程学先, 蒋慧婷 [异构数据源集成实现的研究](#)[期刊论文]-[计算机工程与科学](#) 2008(8)
30. 陈丽娟, 窦迅, 张金望 [基于XML的数据库信息交互](#)[期刊论文]-[福建电脑](#) 2003(11)
31. 孙璐, 张德政 [SOA技术在企业信息资源目录交换系统中的应用](#)[期刊论文]-[有色金属](#) 2009(2)
32. 李颖富 [基于Web服务的配电网设备信息发布系统研究](#)[学位论文]硕士 2005
33. 蔡鲁湘 [基于XML_Schema的地理信息元数据管理系统研究与实现](#)[学位论文]硕士 2005
34. 李爱军 [基于XML的异构数据库交换研究](#)[学位论文]硕士 2006
35. 杜军威 [面向Web的异构数据库交换平台的设计与实现](#)[学位论文]硕士 2005
36. 杨旭 [校园网站内容管理系统的设计与实现](#)[学位论文]硕士 2005
37. 应洁 [异构数据库间通用数据转换模型的研究与实现](#)[学位论文]硕士 2005
38. 李远敏 [基于健康标准电子病历数据抽取模式挖掘研究](#)[学位论文]硕士 2004
39. 张亚峰 [XML技术在基于WEB的产品数据管理中的应用](#)[学位论文]硕士 2006
40. 朱晓华 [基于XML的MEMS材料数据库设计与实现](#)[学位论文]硕士 2005
41. 李莉明 [基于XML的数据转换在电子商务平台E-Hub中的研究与应用](#)[学位论文]硕士 2005
42. 陈强 [城市地理信息系统软件与工程设计](#)[学位论文]博士后 2004
43. 李卫江 [电子政务与空间信息集成的理论及实践](#)[学位论文]博士 2004

引用本文格式: 孙宏伟, 张树生, 周竞涛, 王静 [基于模型驱动的XML与数据库双向映射技术](#)[期刊论文]-[计算机工程与应用](#) 2002(4)