

## 基于 XML 技术的文件的数据库存储

牛方曲<sup>1,2</sup> 朱德海<sup>1</sup> 程昌秀<sup>2</sup> 崔小刚<sup>1</sup>

(1. 中国农业大学 信息与电气工程学院, 北京 100083;  
2. 中国科学院 资源与环境信息系统国家重点实验室, 北京 100101)

**摘要** 为解决 C/S 结构下客户端文档文件的共享问题,采用 XML 技术将不同类型文件转化为 XML 字符串统一存入数据库,读取时只需从数据库中读出 XML 字符并将其恢复为原文件即可。各客户端可对数据进行存取和修改,从而实现了客户端数据的共享操作。操作均在客户端完成,可以充分利用资源,提高系统性能,同时把文件统一存入数据库,简化了基于数据库应用开发的复杂度。本方案在国土资源部土地整理中心的土地整理项目审批系统中成功运用,解决了多种数据的统一管理问题。

**关键词** 可扩展标记语言; 数据; 客户端/服务器; 文档对象模型

**中图分类号** TP 311.11

**文章编号** 1007-4333(2006)03-0101-03

**文献标识码** A

### Storing files in database based on XML technique

Niu Fangqu<sup>1,2</sup>, Zhu Dehai<sup>1</sup>, Cheng Changxiu<sup>2</sup>, Cui Xiaogang<sup>1</sup>

(1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China;  
2. State Key Laboratory of Resources and Environmental Information System, Beijing 100101, China)

**Abstract** Data is now getting more and more various and complex along with the rapid development of network technique. How to store various data efficiently becomes a focus in network and database domains. The C/S structure application based on database often shares data between every client. Some of the data are stored as files, such as Word documents, Excel documents, pictures and so on. In this paper we tried to store data files in database with the XML technique. By this means, data were efficiently shared between every client, and the application was successful.

**Key words** XML; data; C/S; DOM

网络技术的快速发展,电子商务、电子图书、远程教育等的广泛应用所带来的数据的复杂化、多样化、智能化,以及不同用户对同样数据的不同需求,需要复杂的数据库服务,国土资源部土地整理中心土地整理项目管理面临的大量多种数据的管理问题就是一典型例子。如何利用成熟的、已经发展得相当完善的数据库技术对数据进行存储管理并实现有效检索,已经成为当今网络和数据库领域共同关心的问题<sup>[1]</sup>。可扩展标记语言 XML,作为数据传输和交换标准<sup>[1]</sup>,已经在 Web 上得到了广泛应用,在一定程度上发挥了数据库的作用。XML 是互联网联合组织(W3C)创建的一组规范,具有良好的可靠性

与交互性。XML 正在成为 Internet 上数据描述和交换的标准。对于 C/S(客户端/服务器)结构的系统,将 XML 技术用于数据处理,很多情况下能够简化数据库结构和开发工作。为此本研究尝试采用 XML 技术将文件存入服务器端数据库中,实现各客户端数据的共享操作,解决土地整理项目中大量多种数据的统一管理问题。

### 1 C/S 结构系统文档数据的共享问题

C/S 结构的系统中,经常需要将一些文件资源进行共享管理,特别是一些重要的文档资源,如 Word、Excel 文件及图片文件等,使各客户端可以管

收稿日期: 2005-09-03

基金项目: 国家自然科学基金资助项目(40401047)

作者简介: 牛方曲,博士研究生, E-mail: niufq@reis.ac.cn; 朱德海,教授,博士生导师,通讯作者,主要从事地理/国土资源信息系统和农业资源与环境信息系统研究, E-mail: zhudehai@263.net

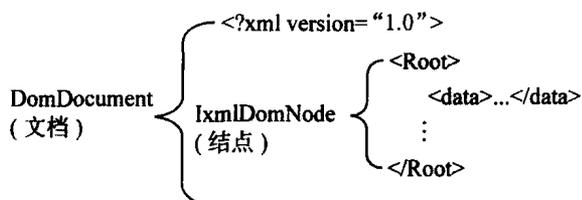
理和使用同一文件资源。这种共享管理不同于局域网中通过网上邻居进行的文件共享管理,而是使各客户端软件之间具有文件共享和互操作的功能。如果将共享文件放在服务器端的磁盘上,由各客户端直接存取,会增加开发和维护的复杂度,并且会给服务器带来不安全性,而且在多用户并发操作情况下其性能低下。将文件以二进制流的形式存入数据库,如 SQL Server 数据库中的 image 类型,可以用来存储文件;但这种方式存在很大不足:写入数据库和读取数据需要复杂的代码,而且不能用通用的 SQL 语句对数据进行操作,如 SQL Server 数据库中的 Image 类型不能用 Insert、Update 等标准 SQL 语法进行修改和添加<sup>[3]</sup>,为数据库操作带来诸多不便。

## 2 以 XML 文档为中介的文件的数据存储

本方案的主要思想为:首先将文件转化为 XML 字符串存入服务器端数据库,读取文件时,从数据库中读出 XML 字符串并将其恢复为原文件。这一系列的操作均在客户端完成,所以充分利用了客户端资源,提高了整个系统性能,同时由于把文件统一存入数据库,简化了基于数据库应用的开发的复杂度。

XML 继承自 SGML 语言,类似于 HTML 语言,但 XML 允许用户自定义标记,具有更大的灵活性,为使用者建立自己的 XML 文档提供了方便。XML 使得在 Web 上能够传输、处理各类复杂的文件<sup>[4]</sup>。目前很多开发工具都提供了处理 XML 文档的功能,这里以 VB 为例对本方案进行详细阐述。

VB6.0 中的文档对象模型 DOM 可用于访问 XML 文档数据。DOM 是互联网联合组织 W3C 推荐的一种技术标准,它提供一组 API 作为应用程序和 XML 文档之间的桥梁<sup>[2]</sup>。根据 DOM 模型,XML 文档对象中包含有结点对象



其中 < Root > < / Root > 与 < data > < / data > 均是结点。

本方案的实现步骤如下:

1) 建立 XML 文档,即将需要共享的文件转换

为 XML 文档。首先以二进制形式读取文件,并将其存放于一个字节类型的数组 ArryByte 中,然后定义一个 XML 文档对象 (DomDocument 类型的对象),并在文档对象下建立结点对象 (XmlNode 类型),将上述字节类型的数组赋给结点。关键代码如下:

```

dim objDom as New Domdocument
dim objNode as XmlNode
...
set objNode = objDom. createElement (" Data ")
...
objNode. nodeTypedValue = Arrybyte
  
```

其中 ArryByte 为已经定义的用于存储二进制文件的字节类型的数组。

这样建立的 XML 文档(只是在内存当中的字符串,并没有存成文件)存储在 objDom 的 xml 属性当中。可以定义一个字符串用于存放该 XML 字符串:

```

dim str as string
str = objDom. xml
  
```

2) XML 字符入库,即将 1) 中建立的 XML 文档存入数据库。由于这里的 XML 文档只是一个字符串,所以只需要用 SQL 语句(如 Insert、addnew 等)即可实现将字符串 str 存入数据库。str 字符串被存储在数据库中的表内,该表中用于存储 str 字符串的字段类型设为字符类型。考虑到 XML 字符串会很长,最好将字段的容量设置大些,如 SQL Server 数据库中可将相应的字段设为 text 类型。

3) XML 字符出库,即将库中的 XML 字符串取出。由于文件是以字符串形式存入数据库的,所以只要相应地定义一个字符串变量 str2,并用 SQL 语句(Select 语句)将库中的 XML 字符串取出赋给 str2 即可。

4) 恢复原文件。再次定义 DomDocument 对象 objDom 和 XmlNode 对象 objNode,并把 str2 赋给 objDom (用 objDom 对象 loadXml () 方法)。因为入库时把文件以字符的形式放入 XML 文档的结点内,所以这里需要取出相应结点的内容。部分代码如下:

```

objDom. loadXml str2      ' 加载 XML 字符串
Set objNode = objDom. selectSingleNode (" Data ") ' 获得结点
ArryByte = objNode. nodeTypedValue      ' 取
  
```

出结点内容

然后以二进制写的方式创建一文件,并将数组的内容存入文件。打开和写入文件代码的关键部分如下。

读取:

```
open FileName for Binary Access Read as File-  
Handler
```

```
Get Filehandler , ,ArrByte
```

```
Close Filehandler
```

写入:

```
open FileName for Binary Access Write as  
FileHandler
```

```
put FileHandler , ,ArrByte
```

```
Close FileHandler
```

其中 FileName 为文件名(包括路径), ArrByte 为字节类型的数组。代码“put FileHandler , ,ArrByte”是将 ArrByte 数组内容写入文件。这样在客户端恢复了原文件。

### 3 应用测试

该方案成功应用于国土资源部土地整理中心(下简称整理中心)的项目审批系统(C/S 结构系统)中。地方政府向整理中心申报土地整理项目时需提供大量材料,这些材料被录入数据库并上报给整理中心供审批。项目中很多材料以电子文档形式存在,如 Word、Excel 文件等,如果将部分申报材料录入数据库,同时将另一部分以文件形式存储在磁盘中,这两部分之间则难以建立关联,而且文档文件如果在一个客户端被修改,其他的客户端(审批人员)无法看见其所做的修改,给整理中心的审批工作带来不便。笔者利用本文 2 中所述的方案,将项目的电子文档和其他数据一并存入数据库,简化了项目的管理,取得了良好的运行效果。

测试环境:操作系统 Windows 2000 Server, 数据库平台 Sql Server 2000, 开发工具为 VB 6.0。

## 4 讨论

本方案不仅可以将 Word、Excel 等文件以 XML 字符的形式存入数据库,而且可以把其他多种类型的文件如图片文件、帮助文件、声音文件等存入数据库。

测试过程中发现 \*.exe(可执行文件)文件不能成功入库,这可能是 Windows 系统禁止以二进制方式读取可执行文件的缘故。

本文中仅概述了方案的实现思想,用户可以根据自己的需要和习惯自行组织代码,例如:可以将代码封装起来做成通用函数或组件,这样可以实现代码的复用,为开发提供方便。

## 5 结论

1) 基于 XML 技术的文件的数据库存储有效地解决了 C/S 结构下各客户端文档文件的共享问题,各客户端可以读取数据库中的数据对其进行修改,然后再次存入数据库。将数据统一存储在数据库中,简化了系统结构。

2) 可以用通用的 SQL 语句对数据进行存取、修改、添加等操作。

3) 可以充分利用数据库系统的并发处理功能。

4) Web 环境(B/S 结构)下同样可以使用本方案将相应数据一并存入数据库统一管理。

## 参 考 文 献

- [1] 胡忠红. 基于 XML 的 Web 搜索技术[J]. 科技情报开发与经济, 2003, 13(1): 88-91
- [2] 陈和平, 张琳. 基于 XML 虚拟后台数据源的 Web 应用[J]. 计算机应用, 2003, 23(1): 108-110
- [3] 雷筱珍. 基于 SQL Server 的 Image 数据类型存取技巧[J]. 福建电脑, 2006(3): 178-179
- [4] 徐永川. SGML、HTML 与 XML 的比较[J]. 情报科学, 2004, 22(1): 100-103