

一种基于XML和Base64编码的数据资源统一存储方法^①

A Method for Data United Storage Based on XML and Base64

简清明 叶晓彤 (四川理工学院 网管中心 四川 自贡 643000)

摘要: 随着信息技术的广泛应用,组织内积累的各种信息资源也日益增加,如何对这些不同格式的数据资源进行统一管理和有效利用就是我们需要面对的问题.提出一种基于XML、数据库和BASE64编码技术的各种数据资源统一存储的方案,实现组织内各种数据资源的统一管理、查询和共享。

关键词: XML base64 编码 数据库 统一存储

随着计算机技术的广泛应用,单位或组织内积累的信息资源如公文,声音、影像、多媒体、图片、工程图纸、电子档案等等也越来越多.这些资源除文本文件外,大多是一些二进制数据,如图像、声音、格式化文档等常见的数据类型.这些数据资源通常以分离的方式存于各个应用系统中,如档案管理系统、CAD系统等并以不同的格式进行存放,对于数据资源的统一管理、共享和远程传输十分不便.本文针对以上问题,提出一种基于XML、数据库和BASE64^[1]编码技术的各种数据资源统一存储的方案,实现组织内数据资源的统一管理和共享。

1 XML技术

XML^[2]是互联网联合组织(W3C)创建的一组规范,适合于松耦合应用程序间以文件格式进行的数据交换.XML的优势在于其处理的数据可以被应用者自行定义的、有语义的标记环绕,可以针对数据表实现无损、安全的数据传输、存储、检索.XML以其良好的可扩展性、平台无关性成为数据表示和交换的新标准.遗憾的是XML是文本型的数据交换结构,对于字符类型的文本交换非常的方便,但对非文本的数据则无能为力,而实际工作中我们常常会处理二进制数据,如图像、声音、格式化文档等,这类数据的高效、无

损传输和表示也是数据交换和存储中需要考虑的问题.这就需要通过某种方式将非文本数据映射为文本数据再进行处理和传输,而Base64编码正是满足这种需要的一种技术。

2 Base64编码

Base64编码是MIME规定的一种编码格式,其基本思想是通过64个ASCII字符对二进制数据进行重新编码组合.Base64编码将连续的3个字节从前向后划分成4个6bit的数值,再从BASE64码表中查出每个6bit数值对应的字符.即将需要转换的数据每三个字节为一组,再将这24位数据按每组6位进行重新划分,在每组的最高2位填充0最终成一个完整的8位字节.如果所要编码的数据的字节数不是3的整数倍,需要在最后一组数据填充1到2个字节的0字节,最后根据BASE64给出的64个基本字符表,查出对应的ASCII码字符.BASE64字符表如下:

ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz0123456789.

3 系统基本结构

整个系统如图1所示,分为两个部分组成:第一步是将数据资源转换成统一格式的xml文档,如为非

① 基金项目:四川省教育厅青年基金(07Zb049)

收稿时间:2009-02-27

文本数据则采用 Base64 编码转换为文本字符串。该过程是可逆的,也就是说可从 xml 文档解码出经过编码的非文本数据。统一格式的 xml 文档的生成使得在各种应用中交换和共享数据资源成为可能,但数据的检索和存储仍然不便。第二步是将 xml 文档通过 DATASTAGE 或其他方式导入大型数据库如 DB2 中,将 XML 中的姓名、性别等基本数据项加载到相应的字段,其中保存数据资源的 data 数据项加载到 CLOB 字段中。CLOB 字段可以存储不超过 2GB 的数据,这对绝大多数的应用包括一般的 DVD 影视都足以满足需要。



图 1 系统基本结构

数据存入数据库以后,借助大型数据库的强大处理能力,资源的存储、检索和共享就变成一件非常简单的事情。各应用程序只需添加一段 BASE64 编码和加码程序即可。BASE64 编码和加码程序也可作为存储过程放在数据库服务器上,以达到对用户透明的效果。

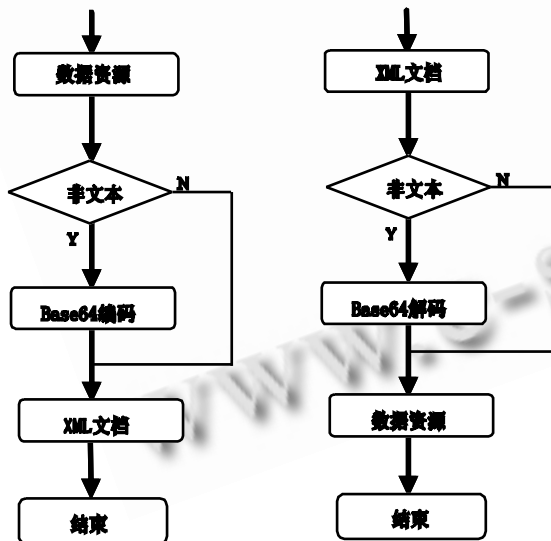


图 2 数据资源的 base64 编码解码流程

```

<?xml version="1.0" encoding="UTF-8" ?>
<Data>
  <Name> 资源名 </Name>
  <type> 资源类型</type>
  <data>/9j/4AAQSkZJRgABAQAA...</dat
    
```

```

a>
  <remark>备注信息</remark>
  ...定义其它信息...
  <Data>
    
```

其中, data 字段保存文本信息或经过 Base64 编码后非文本信息。

4 系统基本结构

可以用各种不同的编程语言实现 base64 编码转换函数。程序的放置可以在服务端,也可在客户端。下面以 java 语言和 SUN 的 API 为例,实现 base64 解码和编码的关键源代码如下:

```

import java.io.*; //Java 核心库,提供全面 IO 接口
import sun.misc.BASE64Decoder; //base64 编码
import sun.misc.BASE64Encoder; //base64 解码

public class Base64
{
  ...
  //以下函数实现文件的 base64 编码
  public static String encode(String fileName) {
    String string = null;
    try { // 建立对象,用于读取文件数据
      InputStream in=new FileInputStream(fileName);
      // in.available()返回文件的字节长度
      byte[] bytes = new byte[in.available()];
      // 将文件中的内容读入到数组中
      in.read(bytes);
      // 进行 BASE64 编码
      string=new BASE64Encoder().Encode(bytes);
      // 关闭文件
      in.close();
    } catch (FileNotFoundException fe) {
      //错误处理
      fe.printStackTrace();
    } catch(IOException ioe) {
    
```

```
        ioe.printStackTrace();
    }
    return string; //返回字符串结果
}
public static void decode(String string,
String fileName) {
    try {
        //进行 BASE64 解码
        byte[] bytes=new BASE64Decoder().
decodeBuffer(string);
        //建立 ByteArrayInputStream 对象
        ByteArrayInputStream in=new ByteArray-
InputStream(bytes);
        //建立缓冲区
        byte[] buffer = new byte[1024];
        //建立输出文件对象
        FileOutputStream out=new FileOutput-
Stream(fileName);
        //初始化参数
        int bytesum = 0;
        int byteread = 0;
        while ((byteread = in.read(buffer)) !=
-1) {
            //将数据写入文件
```

```
        bytesum += byteread;
        out.write(buffer, 0, byteread);
    }
} catch(IOException ioe) {
    //错误处理
    ioe.printStackTrace();
}
}
}
```

5 系统基本结构

通过 xml 和 base64 编码, 将非文本数据转换成文本型数据, 生成统一格式的 xml 文件并导入数据库中, 实现了对多类型资源的统一存储和管理, 为用户有效利用积累的信息资源提供了良好的平台。而统一格式的 xml 文件也可用来用来在不同应用之间进行交互, 为资源共享提供了便利。当然, 转换之间也会消耗一定的系统资源, 这是其不足之处。

参考文献

- 1 Josefsson S, Ed. The Base16, Base32, and Base64 Data Encodings. <http://www.faqs.org/rfcs/rfc3548.html>
- 2 Valentine C.XML Schema 数据库编程指南.北京:电子工业出版社, 2002.1-90.