

基于 Web 的非结构化数据管理方法的研究与实践^①

Research and Practice for Unstructured Data Management Based on Web

文永革 彭声泽 (绵阳师范学院 数学与计算机科学学院 四川绵阳 621000)

摘 要: 结合多媒体网络自主学习平台的开发,分析了 Oracle interMedia 技术框架及工作原理,提出了基于 Web 的非结构化数据在数据库内的便捷管理方法,实现了图像、音视频数据在媒体数据库中的快速存取管理,该方法在基于内容的文本、图像检索方面具有广泛的应用前景。

关键词: Oracle interMedia BLOB 存取管理 格式转换

目前,结构化数据及其管理已得到社会各行业充分的应用,但信息时代越来越多的非结构化数据,如公文、邮件、图片、音视频等数据亟待进一步的管理、利用。据英国 Fulcrum Research 公司的权威数据表明:近 80% 的企业信息数据是没有经过结构化的,可见,重点管理和应用非结构化信息资源已经成为业界一种趋势^[1]。

多媒体数据库的引入对数据管理方法提供了新的思路^[2]。目前,部分 Web 系统采用外部文件的方式来管理,将多媒体文件放在操作系统目录下,系统优点在于反应速度快、数据库及程序设计较简单,但由于所访问文件路径要绝对准确,加上数据管理安全性及数据灾难恢复等方面的原因,更多的应用系统选择采用数据库的管理方式。

本文以《现代教育技术》多媒体网络综合学习平台设计为例,运用 Oracle interMedia 技术,对常用的图像、音视频非结构化类型数据的数据库管理方式进行了应用研究。

1 LOB 数据类型

未结构化的数据是不能被再分为标准的组成部分的数据,通常是指大型对象(Large Object,简称 LOB)类型,在关系型数据库中,多媒体信息等大型对象通常是通过 LOB 型字段来进行存取的^[3]。为了实现对未结构化的数据的存储、管理等操作,Oracle 提出了大型对象

的概念^[4]。

2 Oracle interMedia 组件

interMedia 是 Oracle 8i 开始提出的解决 LOB 存取管理的一种新方案,允许 Oracle 存储、管理、读取图片、音频、视频等非结构化数据,并能对其安全、备份和恢复都能够进行控制管理。

Oracle interMedia 通过类似于 Java 或 C++ 中的对象数据类型,来描述 image、audio、video 数据类型^[5],在数据库中分别被定义为 ORDIImage、ORDAudio、ORDVideo 对象类型。这些对象类型的实例通常元数据、媒体数据及相应操作方法,如对于 image 对象,可使用相应方法方便地进行图形格式转换、压缩、拷贝、截取图形的任何一部份。

2.1 Oracle interMedia 体系结构及工作原理

interMedia 定义了一种数据库服务器、Web 服务器、客户三层框架结构,支持对数据库中传统关系数据及富媒体内容的方便、高效的管理^[6],如下图 1 所示。

凭借 Java 强大的跨平台特性,ORACLE 数据库管理系统在 Oracle 8i 中整合了 Java 虚拟机(JVM),引入了对 Java 的扩展支持功能,利用 Java 来扩展存储程序,在数据库服务器中创建存储过程、触发器等,作为对 PL/SQL 访问机制的有效补充。

① 基金项目:四川绵阳师范学院院级科研项目(071193)

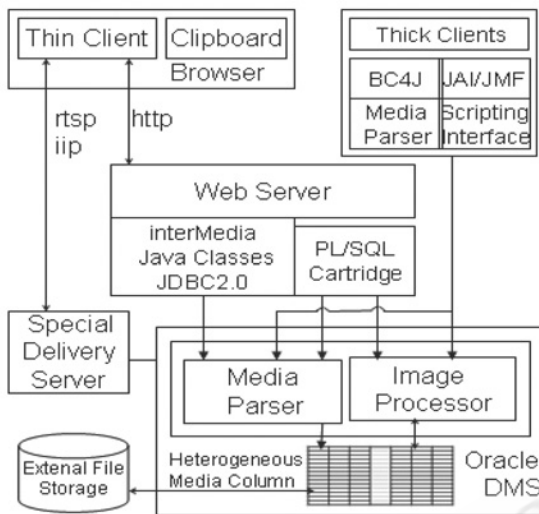


图1 Oracle interMedia 体系结构

在中间层,Web 服务器允许使用 Oracle interMedia Java 类访问 interMedia 对象,对于 Web 应用,Oracle interMedia 同样提供了基于 servlet 及 JSP 的专用包文件,使用 HTTP 协议快速上传下载存储于数据库中的 LOB 数据。

客户端,Oracle interMedia 支持胖、瘦客户的程序设计,对胖客户应用而言,可以使用 JAI、Java 媒体框架 JMF 及商业组件进行可扩展的、多层数据库应用程序的开发。

3 Web 系统非结构化数据管理

多媒体网络学习平台除传统的结构化数据之外,主要使用了图像、音频、视频三种非结构化媒体数据,实现基于 Web 的媒体数据的远程上传、管理及下载等操作。

3.1 Tomcat 容器及配置

随着 Catalina Servlet 引擎的出现,Tomcat 的性能得到提升,成为 SUN 公司推荐的 Servlet/JSP 容器,同时,Tomcat 是 Java Servlet 和 JavaServer Pages 技术的标准实现,与 Servlet API 和 JSP 完全兼容。系统选择 Tomcat 作为 WEB 服务器。

3.2 BLOB 数据的 Web 管理实现

Oracle 提供的 ORDImage、ORDAudio、ORDVideo 对象的访问操作方式大致相同,此处以 ORDImage 对象

操作为例,实现瘦客户的多媒体数据文件的上传、管理、回传等操作。

3.2.1 媒体文件上传

利用 Ordhttp 自动获取媒体文件信息,Ordhttp 是 Oracle 提供支持 Servlets 和 JSPs 的 interMedia Java 包,包括 OrdHttpUploadFile 和 OrdHttpUploadFormData 两个子类。

在 JSP 程序中,获取客户端上传文件的实例对象 uploadImage:

```
<jsp:useBean id="formData" scope="page"
class="oracle.ord.im.OrdHttpUploadFormData" />
```

```
OrdHttpUploadFile uploadImage = null;
```

```
uploadImage = formData.getFileParameter("
media");
```

利用 OrdHttpUploadFile 类提供的方法,获取 ORDImage 对象属性:

```
filename = uploadImage.getSimpleFileName();
//文件名
```

```
filetype = uploadImage.getMimeType(); //文件
类型
```

```
filelength = uploadImage.getContentLength(); //
文件长度
```

在 JavaBean 中实现获取 ORDImage 对象引用:

```
image = (ORDImage)rs.getCustomDatum(1,
ORDImage.getFactory());
```

```
uploadImage.loadImage(image); //上传 OR-
DImage 对象
```

3.2.2 媒体格式转换及尺寸缩放

图像文件所包括的最小信息包括尺寸、像素位数、压缩类型及内容格式^[7],使用 Oracle interMedia Java 包提供的 ORDImage 类相关方法,在 JavaBean 中可以方便地实现图片格式、尺寸的管理:

图片格式的转换:

```
if (image.getMimeType().indexOf("bmp") >
0)
```

```
image.process("fileFormat=" + getPreferred-
Format(image.getContentFormat()));
```

```
//bmp 转 gif 或 jpg 格式
```

```

if( numDigits > 0 ) // numDigits 为图片色彩
位数
if( Integer.parseInt( contentFormat. substring( 0 ,
numDigits ) ) > 8 )
return " JFIF" ;
//色彩位数大于 8 转换为 JPEG 格式
return " GIFF" ;
//否则 转换为 GIF 格式

```

图片尺寸的缩放：

```

image. processCopy( " maxScale = 300 ,300" ,
thumb );
//thumb 为目标对象

```

3.2.3 从数据库中访问、下载媒体对象

JavaBean 在获取 ORDImage 对象后 ,通过 Ordhttp 包中 OrdHttpJspResponseHandler 子类提供的 sendImage()方法来完成向浏览器中的发送。

```

<jsp :useBean id = " getsbean" class = " mmes.
GetResultSet" scope = " page" / >
<jsp :useBean id = " handler" scope = " page"
class =
" oracle. ord. im. OrdHttpJspResponseHandler" / >
.....
handler. sendImage( getsbean. getImage( ) );

```

在学习平台的测试环节 ,图片可根据指定的显示范围动态调整大小 ,显示试卷上 ,并可根据测试者要求进行放大观看 ,如图 2 所示。

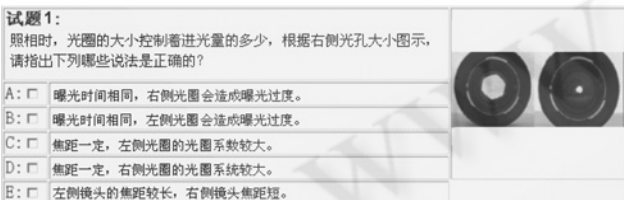


图 2 OrdImage 对象在 IE 中的显示

对于音、视频大对象数据类型 ,在发送到浏览器时 ,可以判定返回对象的媒体类型是否是 audio/x - pn

- realaudio、application/x - pn - realmedia 等 ,据此动态嵌入相应的播放器 ,保证获取媒体文件的正确播放。

4 小结

Oracle interMedia 多媒体组件为开发非结构化数据管理系统提供了强大的功能及便捷的访问接口。本系统应用其存取、管理技术实现了图像等非结构化数据的数据库管理 ,在多媒体网络课程学习、测试中成功地得到了应用 ,同时 ,为数据的安全管理、事务处理及灾难恢复提供了一定的保障 ,但是 ,在另一方面 ,对于音视频文件 ,尽管 Oracle 支持的 4GB 单个存储对象 ,考虑到数据库在频繁访问时的系统工作效率 ,应尽量考虑较小的音视频片段。

参考文献

- 1 中国计算机报. TRS: 专注非结构化数据管理[EB/OL]. <http://www.ciw.com.cn/News/info/2006-06-01/.5389.shtml>, 2006-06-01.
- 2 文燕平. Oracle 数据库管理视频信息. 现代图书情报技术, 2004, 106(1): 50 - 52.
- 3 谈竹贤, 王毅, 赵景亮. Oracle 9i PL/SQL 从入门到精通. 北京: 中国水利水电出版社, 2002. 484 - 485.
- 4 马传松, 邓静, 李振坤. 在 ORACLE 数据库中如何对多媒体数据对象进行处理. 广东自动化与信息工程, 2003(3): 40 - 42.
- 5 林宏基, 王瑞章. GIS 多数据源与 Oracle 数据库的集成. 福州大学学报(自然科学版), 2002, 30(2): 185 - 187.
- 6 Oracle: Oracle interMedia User's Guide and Reference. 英文版[DB]. 2001.
- 7 Oracle White Paper: Oracle Multimedia: Managing Multimedia Content. 英文版[DB]. 2007.