使用XML实现动物疾病关系数据库的语义映射

周 全,李 旸
(安徽农业大学 计算机系,安徽 合肥 230036)

摘 要:关系数据库是当今农业信息存储的主要形式。随着Web技术的发展,信息检索越来越复杂,关系数据库需要更好被Web理解,需要更多语义上下文支持;使用XML格式文档来表达和存储数据的目的就是要解决这方面问题。文中以常见农业动物疾病信息数据库为例,通过比较两种数据存储表示形式找出XML文档结构的优势并使用Java语言设计映射算法,把现有禽类疾病关系数据库部分内容映射为XML数据形式,证明了该存储方式的优越性,为今后农业动植物疾病的语义网建设打下基础。

关键词:XML;关系数据库;数据映射;农业动物疾病
中图分类号:TP39 **文献标识码**:A **文章编号**:1673-629X(2010)09-0243-03

Implementation Semantics Mapping of Animal Diseases Relational Database By Using XML

ZHOU Quan, LI Yang
(Computer Department of Anhui Agricultural University, Hefei 230036, China)

Abstract: Relational database is now the main form of agricultural information stored. But complexity of information retrieval becomes more complex with the development of Web, relational database need to be better understood by Web, which needs more semantic supports. To solve the problem, XML format can be used to express and store data documents. This paper takes common agriculture animal disease information database for example, by comparing the two kinds of data forms to find out the advantages of XML document structure and use the Java design mapping algorithm, maps some parts of the contents of the existing poultry disease relational database to the XML data format, shows the superiority of this way, lay the foundation for the future semantic Web construction of agriculture animal and vegetation diseases.

Key words: XML; relational database; data mapping; agricultural animal disease

1 现在农业关系数据库的不足

我国地域辽阔、农产品种类丰富,互联网上各种农业信息频繁交流传递。Web技术及其应用的快速发展和现代化农村建设的需要,使农业信息结构发生改变[1]。这些信息基本上分为两大类:结构化数据信息、非结构化数据信息。如新城疫这一家禽常见病,结构化信息是指可以用二维表结构表达的疾病名称、病原、发病机理等,而该病的图片、视频和文档类的信息等很难或无法用数据库二维逻辑表表达的称为非结构化信息,结构化数据实际上是非结构化数据的特例;另外还有一种半结构化数据,具有不规则、多变的结构特征,在当今的互联网上也广泛地存在[2]。

目前农业领域用到的数据库大部分是关系型数据库(Relational Database,RDB),其最大缺点在于很难直接构造与具体应用相关的信息类型的表达能力。通常,数据库之间用于信息交换的文件格式都非常简单(如文本文件即可),例如每行一个记录,记录的域之间使用定界符(如分号等)隔开即可。然而这种方法还不能达到面向对象信息的需要;因为通常对象可能还会拥有内部结构,而且对象之间可能还有特定的联系;因为第一范式的原因,现在的RDB要求所有的数据必须为整数、实数、字符串等简单类型,这样复杂的数据类型就必须要被分解,这一工程耗时耗力,而且被分解的数据结构不能直接表示数据。比如现在国家一级的大型农业信息数据库以Oracle为主,而各县乡和科研院所等基层单位使用的数据库因各自需求和经费能力而不同,这些数据库的数据信息彼此异构,很难达到无缝集成共享[3]。

收稿日期:2009-12-31;**修回日期**:2010-03-27
基金项目:安徽省信息产业发展计划(2008-01-25)
作者简介:周 全(1985-),男,安徽合肥人,硕士研究生,研究方向为计算机网络可靠性研究;李 旸,博士,副教授,硕士生导师,研究方向为计算机网络可靠性研究。

其次,关系数据库本身存在语义表述不清的缺点。关系数据库中的记录是属性孤立的,缺乏规范化的上下文结构[4],例如,用搜索引擎在关系数据库中查询“新城疫”,出现的结果可能是该病本身的详实信息,也有可能是与该病相关的文章、书籍或相关会议的内容,这与查询者的本意大相径庭,这就是因为该字段信息不具备基本的上下文信息,使得 Web 无法判断该字段属于哪些语义环境,所以结果中出现很多无关信息,查询者再从这些信息中一一筛选自己需要的信息则十分费力。据 Google 统计,平均每次查找某特定信息所浏览的网页有一半是无用的。

此外,关系数据库在复杂信息的查询时可能要处理大量的表和复杂的连接运算等过程,因此要求用户对 SQL 语句十分熟悉,如果专门开发某应用程序进行查询,则又会造成大量维护相关应用软件的开销;由于与编程语言所提供的数据类型不一致,也导致关系数据库对环境有着较高的要求,不同环境的转换成本高且困难。

随着农业基础环境的不断发展和完善,统一数据的表示格式已成为最主要的问题。可扩展标记语言(XML)以其结构化、可扩展性、灵活性和可验证性成为数据描述和传输的基本方法,因此,将描述农业信息的数据转换成 XML 文档是实现农业信息集成的一种重要手段和方法。

2 XML 数据的优势特点

从某种意义上讲,一个 XML 文档就是一个数据库或其中的一张数据表;XML 是 Web 上定义数据的通用语言,并已逐渐成为数据库信息交换的重要工具之一[5]。XML 允许为指定的应用程序创建一致的数据格式,也是服务器间传递数据的理想格式[6]。概括来说,XML 格式表示的数据源有以下优势:

1)唯一性:XML 格式的数据因为使用丰富的标签来定义上下文语义而被唯一标记,能用于更精确的检索。如上文“新城疫”的例子。

2)动态的层次结构更新:易于实现不同的粒度更新。XML 只把有变化发生的元素从服务器传给客户,不必重传整个结构化数据[7]。

3)自描述性:XML 通常包含一个如 DTD 的文档类型声明,不仅人能读懂 XML 文档,而且计算机也能处理;XML 文档中的数据可以被任何能够对 XML 数据进行解析的应用所提取、分析和处理,并以所需格式显示;XML 表示数据的方式真正做到了独立于应用系统,这样就赋予数据重用性,可以更好地实现数据的共享和跨平台操作[8]。

4)数据集成性:从离散的 XML 数据源集成数据,在多种不兼容的数据库中检索是难以实现的,但 XML 格式的数据源可以很方便地被集成到中间层服务器上,这些数据还可以根据需要进行进一步的集成、处理和分发。

5)易读性:XML 是标准通用的描述语言,能够被程序操作解析。而对于关系数据库,在不同应用里只有用配套的数据源应用程序才能访问(如. NET 和 ADO. NET,JSP 和 JDBC 等),但在 XML 中,只要程序能够解析 XML 结构,就可以读取关系数据表中的信息。

XML 以其方便表达异构数据和强大的语义能力对关系数据库不足之处做了很好的弥补,使用户在可以处理复杂异构农业数据的同时,面对相似或相同的记录信息时也可以对其语义准确定位[9]。例如,在关系数据库中查询 43～44℃,不论是人还是机器都无法识别其所代表的意思,比如鸡类患新城疫、猪在发生猪丹毒时都可能出现这样的体温,而如果用 XML 来表达,就会准确定位这是农业数据库下动物疾病里常见家禽类疾病的新城疫发病初期的危险体温,如图 1 所示。

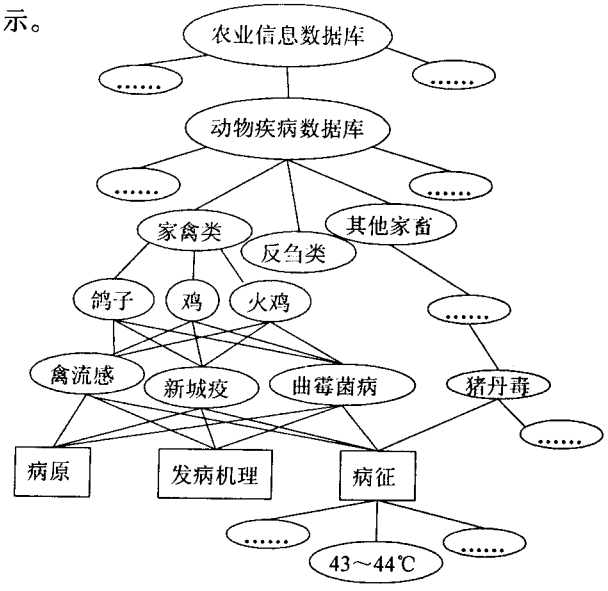

图 1 农业数据库结构示例

以上分析说明了 XML 相对关系数据库在某些方面有着不可替代的优势,那么如何把现有存储在关系数据库中的数据映射为 XML 数据表达是相关研究中首要解决的问题。

3 关系数据库到 XML 的映射算法

从图 2 可以看出 XML 到关系数据库之间存在一种映射关系。但是目前信息的主要存储方式还是在关系数据库中,用 XML 格式来存储数据仍处在起步试

验阶段,离真正投入实际使用还有很长的路要走。我国农业类数据也不例外,如果用人工把这些数据变成 XML 文档,其工作量浩大且效率低下。因此尽管通过以上分析可以看出 XML 格式文档具有明显优势,但要通过关系数据库操作实现 XML 的事务管理,首要解决的问题是要把已有关系数据库的内容映射为 XML 文档[10],这种映射算法的实现方法较多,以下给出其中一种描述。

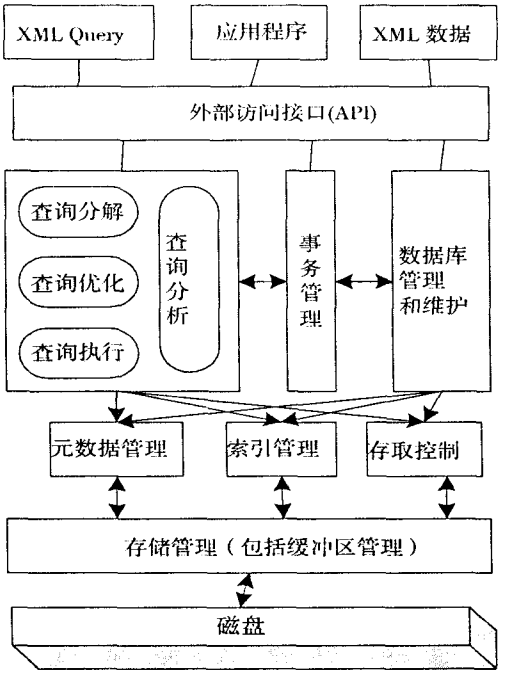

图 2　XML 数据库体系结构

先给出简单的将关系数据库字段转换为 XML 格式文档的方法,核心就是把字段和属性值都存于合适的数据结构中,再将其循环调用,在调用时自动匹配上相应标签,以至分辨不同层次结构。

```
//List 为相应的链表数据结构
List list1 = 获取 RDB 的字段名称;
List list2 = 获取 RDB 的属性值;
for ( int i = 0;i< = list1.size();i + + )
{
for (int j = 0;j< list2.size();j + + )
{
String tag = '<' + list1.get(i) + '>'
+ list2.get(j) + '</' + list1.get(i) + '>';
Record();//此函数记录或显示此时的标签和内容
if(j = = list2.size() - 1)//如果 j 已经循环结束则置 0,重新循环
{
j = 0;
break;
}
break;
}
```

以上算法实现了关系数据库到 XML 文件的映射,优点在于可以清晰方便地表达一些非结构化数据,并把原本彼此孤立的字段信息都用严谨的上下文语义联系了起来,使得所有数据都具有唯一性,提高了信息检索效率。

4　试验及结果展示

文中在农业动物疾病关系型数据库的基础上,以鸡类疾病表为例,在 J2EE 平台 Struts + Hibernate + Spring 框架下实现农业动物疾病关系型数据库映射到 XML 文件[11],试验环境:

主机配置:CPU:1.6GHz,1G 内存

操作系统:Windows XP

IDE:MyEclipse6.5 + Tomcat5.0 + jdk1.4

数据库:MySQL5.0

开发框架:Struts1.2 + Hibernate3.0 + Spring2.5

作为描述 XML 文档信息结构数据模型(即词汇表),DTD 是必不可少的,以上数据对应的 DTD 文档结构:

```
〈? XML version = "1.0" encoding = "GB2312" ?〉
〈! DOCTYPE animaldisease[
〈! ELEMENT poultrydisease (Chicken * , Duck * , Goose * , Pigeon * )〉
〈! ELEMENT Chicken (diseasename, etiology, pathogenesis, picture, symptom, prevention)〉
〈! ELEMENT diseasename (#PCDATA)〉
〈! ELEMENT etiology (#PCDATA)〉
〈! ATTLIST etiology viruscarrier CDATA #REQUIRED〉
〈! ATTLIST etiology disinfection CDATA #REQUIRED〉
〈! ELEMENT pathogenesis (#PCDATA)〉
〈! ELEMENT picture (#IMPLIED)〉
〈! ELEMENT symptom (#PCDATA)〉
〈! ATTLIST symptom mortality CDATA #REQUIRED〉
〈! ATTLIST symptom incubation_ period CDATA # REQUIRED〉
〈! ATTLIST symptom symptom_ type CDATA #REQUIRED〉
〈! ELEMENT infect_ animal (#PCDATA)〉
〈! ELEMENT prevention (#PCDATA)〉
]〉
```

数据库表结构就不再赘述。

工程中主要包及其功能如下:

action:用于存放业务处理类;Tool:存放工具类,如分页等重用性代码;Dao:数据访问对象,用于访问持久层对象;Form:持久化对象。

过程描述:通过输入页面进行疾病相关信息的输

(下转第 249 页)

handle,int Type, int * pVideoCode, int CodeLen),参数含义:handle h[in],视频解码器的句柄,int Type[in],帧类型,int * pVideoCode[in],待解码视频流指针,int CodeLen[in],待解码视频流长度;获得解码后的视频图像函数 int GetVideo(int handle, int * pOut, int rowspace) 参数含义:int handle[in],视频解码器的句柄,int * pOut[out],指向 BYTE 类型的视频图像缓冲区指针,int rowspace[in]输出视频画面前后两行相隔的差距。

5 结束语

该设计采用了高效的 H.264 编码技术和实时媒体传输技术,能满足视频监控系统的需求,且可扩展性强。本系统采用特定嵌入式场合的专用 Linux 操作系统,裁减后,放在容量只有几百 k 字节或几兆字节的 Flash 芯片中。下一步将加入智能化模块,使系统不断提高和完善。

参考文献:

[1] 苏光大 . 微机图象处理系统[M].北京:清华大学出版社,2000.

[2] 郭宝龙,倪　伟,闫允一 . 通信中的视频信号处理[M].北京:电子工业出版社,2007.

[3] JVT of ISO/IEC MPEG and the ITU - T JTC: ISO/IEC 14496 - 10:2005 Information technology - Coding of audio - visual objects - Part 10: Advanced Video Coding[S]. 2005.

[4] Ostermann J,Bormans J,List P,et al. Video coding with H.264/AVC: tools,performance, and complexity[J]. IEEE Circuit and Systems magazine,2004,41:7 - 28.

[5] 毕厚杰 . 新一代视频压缩编码标准——H.264/AVC[M].北京:人民邮电出版社,2005.

[6] Texas Instruments. TMS320C64x DSP Library Programmer's Reference[M].[s.l.]:[s.n.],2004.

[7] Texas Instruments. TMS320C6000 CPU and Instruction Set Reference Guide[M].[s.l.]:[s.n.],2004.

[8] Texas Instruments. TMS320C6000 Assembly Language Tools User's Guide[M].[s.l.]:[s.n.],2004.

[9] 彭启踪,管　庆.DSP 集成开发环境——CCS 及 DSP/BIOS 的原理与应用[M].北京:电子工业出版社,2004.

[10] Texas Instruments Incorporated. TMS320C6000 系列 DSP 编程工具与指南[M].田黎育,何佩琨,朱梦宇 编译.北京:清华大学出版社,2006.

[11] 魏本杰,刘明业,章晓莉.二维 DCT 算法及其优化的 VLSI 设计[J].计算机工程,2006,32(2):16 - 18.

[12] 陈　明,梁兴东,吴一戎.基于 H.264 的嵌入式无线视频监控系统[J].微计算机信息,2008,32(5 - 2):10 - 12.

(上接第 245 页)

入(名称、病原、发病机理等),可实现相关疾病信息的查询,并生成与数据库字段对应的 XML 格式文件。

5 结束语

文中通过对传统关系型数据库和 XML 格式数据特点的比较,在 J2EE 平台上借助成熟的开发框架,将现有的农业动物疾病关系数据库映射为 XML 形式的结构化文档,在此基础上,可以对 XML 文档做更精确的信息转换和查询[12],实现关系数据库全部内容映射为 XML 结构文档,为今后该方向专家系统和语义网的建立打下良好基础。

参考文献:

[1] Ronald B. XML and Databases[DB/OL]. 2009 - 05 - 11. http://www.rpbourret.com/xmldbms.

[2] Widom J. Data Management for XML: Research Directions [DB/OL]. 2009 - 05 - 11. http://www - db.stanford.edu/~widom.

[3] 周法国,王映龙.非结构化信息抽取关键技术研究探讨[J].计算机工程与应用,2009(14):1 - 6.

[4] 李志辉.XML Schema 语义约束在关系数据库中的实现[J].计算机与现代化,2009(10):33 - 37.

[5] Kong L B,Tang S W,Yang D Q,et al. Querying Techniques for XML Data[J]. Journal of Software,2007,18(6):1400 - 1418.

[6] 周爱武,李孙长,程　博,等.XML 数据库的研究与应用[J].计算机技术与发展,2009,19(9):218 - 221.

[7] 强保华,潘家志,余建桥.从关系数据库中生成 XML 数据源的研究[J].计算机科学,2002,29(5):70 - 71.

[8] He Z Y,Li J Z,Wang C K. A data model for XML database [J]. Journal of Software,2006,17(4):759 - 769.

[9] Goldman R,McHugh J,Widom J. From Semistructured data to XML [C]//Proc of the 2nd Workshop on Web and Databases.[s.l.]:[s.n.],1999:25 - 30.

[10] 薛星云.探索 XML 模式与数据库模式之间的映射[J].福建电脑,2009(9):176 - 176.

[11] 肖辉辉,段艳明,兰小机.基于 Hibernate 的 XML 数据存储方法[J].计算机系统应用,2009(10):189 - 192.

[12] 周　健,孙丽艳.面向对象 XML 的存储模式的研究[J].计算机技术与发展,2009,19(3):114 - 117.