

文章编号:1009-038X(2001)05-0522-04

基于XML的汽车零部件网站数据 存储与查询的实现

李志华, 施明辉, 孙荣胜

(江南大学信息工程学院, 江苏无锡 214036)

摘要: 以一个汽车零部件网站为例,介绍了基于XML的数据存储,结合关系数据库、MSXML提供的JAVA类等工具,实现了XML信息查询的技术。

关键词: 标记语言;文档类型声明;汽车零部件;存储;查询系统

中图分类号: TP 311.113

文献标识码: A

The Realization of Storage and Query for Parts of Cars on Internet Based on XML

LI Zhi-hua, SHI Ming-hui, SUN Rong-sheng

(School of Information Engineering, Southern Yangtze University, Wuxi 214036, China)

Abstract: A new method to solve the problem of the data which characterized with layer was found in the paper, and the method which was developed from the traditional way combines XML technology with DBMS technology was also obtained. This paper also introduced the research for combining XML with organization, query and dissemination of the parts of cars on Internet.

Key words: XML; DTD; parts of cars; storage; query system

XML(Extensible Markup Language 可扩展标记语言)同HTML一样,是SGML(标准通用标记语言)的一个子集,定义了结构化表达数据的标准格式。在表达复杂的产品数据文档时具有很大的优势。在存储信息方面,其数据的结构性虽然不及数据库,但比数据库有更好的描述能力,比较适合用于描述半结构数据,并且XML文档是自描述的,结构和内容共存,与显示实现了真正地分开。

近年来,基于Web数据库的XML应用,大多是通过开发中间程序将传统的数据库中的数据转化为XML文档,用XML-QL进行查询,并将结果

用XML文档显示给用户。这一方式存在以下不足:①当数据量相当大时,转换后的xml文档将占用相当大的存储空间,造成存储空间的重复浪费;②对具体数据库的基于脚本级的静态转换,实际意义不是很大,而要实现针对Internet上的异数据库的动态转换,中间程序将相当复杂,难以实现。

本系统通过把DBMS和XML相结合,借助传统DBMS在数据库检索方面的优势和成熟技术,用SQL查询后台数据库取得部件XML文档存储位置、零件编号等信息,根据零件编号“Num”关键字,从文件服务器中的XML文档中取得相应汽车零部件

收稿日期:2001-06-04; 修订日期:2001-09-13.

基金项目:江苏省计算机信息处理技术重点实验室项目(K99088)资助课题。

作者简介:李志华(1969-),男(土家族),湖南保靖人,计算机应用硕士研究生,讲师。

件的详细描述,并用 XML 将结果显示给用户,有效地避免了上述问题。

1 基于 Internet 汽车零部件网站的体系结构

本系统的建设主要是为了面向零部件生产厂家、中间商、仓储商、私人用户在生产、经营汽车零部件和汽车维护时,能对整个行业产品的行情有一个较全面地了解。

整个系统的设计是通过 DBMS 与 XML 数据相结合的方法实现的,其中查询子系统的设计与平台无关。整个系统的核心是系统管理程序,它是通过继承微软 XML DOM(document object model,文档对象模型)分析器 MSXML 中的 JAVA 类编写而成的一组 Java 程序,主要负责获取查询关键字并协调汽车零部件的 XML 元数据与后台数据库之间的联系,并访问汽车零部件 XML 元数据,数据库的各表中主要是为了方便用户查询而保留的一些查询关键字和 XML 元数据的存储信息。Parser 是 XML 语法分析器,主要是对 XML 文档和 DTD 进行解析和合法性检验。系统中应用现成的 Java Parser,如 IBM 的 XML 4J 或 SUN 公司的 Project X,它们都符合 SAX1.0 和 DOM1.0 标准。整个系统的体系结构如图 1 所示。

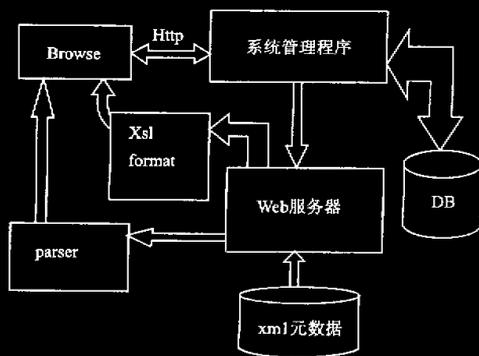


图 1 系统的体系结构

Fig.1 The structure of the system

2 汽车零部件数据的存储

2.1 汽车零部件数据的数据结构

为了实现高效率的网上查询,必须很好地解决汽车零部件数据的存储问题,设计过程中,通过对汽车零部件的分析,得出如下数据结构,如图 2 所示。

很显然这是一棵树,系结点的子结点是部件结点,部件结点的子结点是零件结点,零件结点的叶结点即为该零件的描述信息,也就是用户需要的信息。

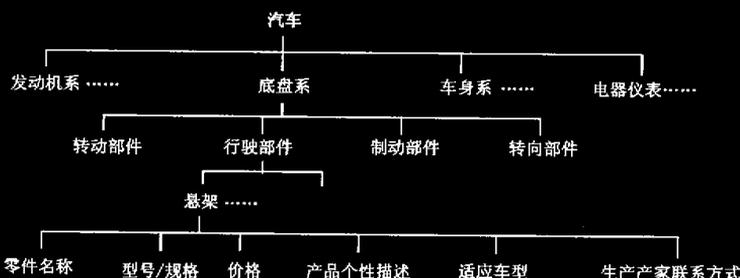


Fig.2 The data structure for the parts of cars

2.2 数据结构到 XML 的文档映射

2.2.1 DTD 的设计 为了提高汽车零部件信息的通用性,系统利用 XML 对汽车零部件信息进行统一描述。用 XML 定义了适合远程查询的 DTD,以部件的 DTD(parts.dtd)为例,具体定义如下:

```

<! DOCTYPE DOCUMENT
[(! ELEMENT 部件(产品 * )
(! ATTLIST 部件 id CDATA #REQUIRED)
(! ELEMENT 产品(零件名 + |型号规格|价
格|产品个性描述?|适应车型|生产产家))

```

```

(! ATTLIST 产品 id #REQUIRED)
(! ELEMENMNT 零件名(#PCDATA))
(! ELEMENMNT 型号规格(#PCDATA))
(! ELEMENMNT 价格(#PCDATA))
(! ELEMENMNT 产品个性描述(#PCDA-
TA))
(! ELEMENMNT 适应车型(车型 *))
(! ELEMENMNT 车型(#PCDATA))
(! ELEMENMNT 生产产家(厂家 *))
(! ELEMENMNT 厂家(名称|地址|邮编|联系

```

电话|传真|电子邮件|网站?)
 <! ELEMEMENT 名称(#PCDATA)
 <! ELEMEMENT 地址(#PCDATA)
 <! ELEMEMENT 邮编(#PCDATA)
 <! ELEMEMENT 联系电话(#PCDATA)
 <! ELEMEMENT 传真(#PCDATA)
 <! ELEMEMENT 电子邮件(#PCDATA)
 <! ELEMEMENT 网站(#PCDATA)]

<产品个性描述>铸造</产品个性描述>
 <适应车型>
 <车型>奥迪</车型>
 <车型>桑塔纳</车型>
 <车型>夏利</车型>
 </适应本型>
 <生产产家>
 <厂家>
 <名称>无锡开源集团</名称>
 <联系电话>5807602</联系电话>
 <传真>5287122</传真>
 <电子邮件>XJ@263.net</电子邮件>
 <地址>江苏.无锡</地址>
 <邮编>214031</邮编>
 <网址>http://www.autopart.org</网址>
 </厂家>
 </生产厂家>
 </产品>
 <产品 id="002">——</产品>
 </行驶部件>

DTD的引用有两种方式:内部 DTD 和外部 DTD,在本系统中使用外部 DTD.

2.2.2 系、部件、零件结点到 XML 文档的映射

XML 可以灵活地定义具有确切含义的元素及其属性且结构嵌套.从汽车零部件的数据结构图分析可知,用 XML 标记语言能方便地对此加以描述.

每一个部件结点由若干个零件子结点组成.即每一个部件 XML 文档由若干个“<产品>”tag 组成.每一对“<产品></产品>”tag(即一棵“产品”子树)对应一个零件结点.其中的 xml 描述的就是该零件的详细信息.以“行驶部件”的 XML 文档的典型结构为例,零件“悬架”是它的其中的第一棵“产品”子树,XML 文档如下:

```
<? xml versoin = "1.0" encoding = "GB2312" ?>
<! DOCTYPE 行驶部件 SYSTEM "parts.dtd">
<行驶部件 id="22000">
<产品 id="001">
<零件名>悬架</零件名>
<型号规格>独立悬架</型号规格>
<价格>89.00</价格>
```

其中部件的“id”号在系统中是唯一的,前两位分别对应部件所屬系和系内的部件编号.后3位限定零件编号的长度,不超过3位.

2.2.3 部件 XML 元数据到关系数据库的映射

为了实现通过数据库查询,然后访问部件 xml 文档,取得零件结点的详细信息.必须实现数据表到各部件 xml 文档的存储位置和零件结点的映射.实现映射的数据表条目的数据格式及示例如表 1 所示.

表 1 数据表条目格式、示例

Tab.1 The format of table entries and examples

部件编号	零件编号	零件名称	名称 1	名称 2	—	文件名
22000	001	悬架	悬架总成	—	—	www.autopart.org/xs.xml

其中,名称 1、名称 2、名称 3 是该产品的可能别名或商业名称,以方便模糊查询.

3 网上汽车零部件查询的实现

3.1 查询功能概述

为了方便用户查询,本系统提供以下 3 种查询方式:

- 1) 按汽车的体系结构进行查询
如:从系(底盘系)→ 部件(行驶部件)→ 零件(悬架);

- 2) 按零件所适应品牌车型和零件名称的方式查询
如:桑塔纳 2000 的悬架;

- 3) 按零件的生产产家和零件名称进行查询
如:无锡开源集团的扭振减振器;

3 种不同的查询方式,为了设计上的方便,最后将以相同的格式提交给客户,都包括:零件名称,型号/规格,价格,产品个性描述,适应车型,生产产家联系方式等信息.

3.2 查询系统的实现:

本系统实现了在 Internet 上对汽车零部件的模

糊查询。在图型界面设计中,主要采用按钮控制。从图型界面的下拉列表框中选择适当的查询方式。在模糊查询编辑框中输入查询关键字,然后单击提交按钮,符合条件的所有不同产家生产的汽车零件产品的信息以及联系方式以表格的形式显示出来。

对汽车零部件的查询,概括起来主要包括以下2个步骤:

第1步,访问后台数据库,找到零件编号“Num”和对应部件XML文档的存储位置;

第2步,从部件的XML文档中取出符合条件的零件结点的XML描述(即零件叶结点);

最常用的“品牌车型和零件名称”的组合查询方式为例,具体讲述查询的实现过程。

3.2.1 查询后台关系数据库,获得零件编号“Num”和对应部件的XML文档的存储位置

建立数据库与JDBC的连接。JDBC是JAVA语言的数据库访问API,JDBC提供了特殊的驱动程序JDBC-ODBC桥。该驱动程序支持JDBC通过现有的ODBC驱动程序访问其数据库系统。本系统中,由于微软的SQL Server 7.0支持ODBC,因此要建立与ODBC数据库的连接。方法:先加载一个JDBC-ODBC桥驱动程序,然后指定JDBC URL,使它指向网上的资源,格式URL url="jdbc:odbc:CN-CARDB";其中CNCARDB为数据源数据库名,通过JDBC API提供的一系列的类接口对数据库操作。对数据库的一个典型的操作如下:

```
Connection db = DriverManager.getConnection
(url);//连接数据库
```

```
Statement st = db.createStatement();//创建一个对象 st
```

```
String sql-str = "select * from LBJYSB where
零件名称 = selectbox.getText() and (零件名称 like
%edit1.getText()% or 名称1 = edit1.getText() or
名称2 = edit1.getText() or 名称3 = edit1.getText
())" ;//定义SQL语句
```

```
Result rs = st.executeQuery (sql-str);//执行
SQL语句,并返回结果集
```

“LBJYSB”为零部件映射表。通过把SQL语句传给executeQuery构造器执行数据库操作。构造器返回一个结果集(ResultSet)对象,当执行SQL查询后,通过处理结果集对象的内容,来获得零件的编号“Num”和对应的部件XML文档的存储位置。其中,部件xml文档的文件名存放在字符串变量“filenamepath”中。如“悬架”的零件编号为“001”,对应的部件xml文档的文件名为:“www.autopart.org/xs.

xml”。

3.2.2 访问文件服务器上的部件XML文档,获取零件的XML描述

由于XML文档是以文件的形式存储在文件服务器上,所以在系统中,对汽车零部件的查询,最后得来的查询信息主要是通过访问文件服务器上部件XML文档中的“产品”子树来实现的。

方法:在系统管理程序中,导入MSXML分析器中的相关Java类,并定义自己的类:ReadXml和ViewElement等。

类ReadXml是MSXML的Java类库中“Com.ms.xml.om.Document”类的一个子类,在该类中,首先,创建一个新的对象“Document xmlfile = new.Document()”,该对象中,主要包括Load(url)方法,url指定元素所在的XML文档的路径,并加载该文档。如零件“悬架”对应的部件xml文档为:www.autopart.org/xs.xml。另一个重要的方法,即通过重载applet类的action方法来对具体的单击事件进行处理。

Showproducts(int num, Element root1)方法是ViewElement类的又一个主要方法,该方法是在找到的部件XML文档中,取得第“Num”棵“产品”子树的全部信息。该方法的主要逻辑结构如下:

```
Showproducts(int num, Element root1)
|
Element elem1, elem2;
ElementEnumeration Enum;
Enum = new ElementEnumeration(root1);
For (int I = 0; I <= num; I + +)
    elem1 = (Element)enum.nextElement();
ElementEnumeration Enum2;
Enum2 = new ElementEnumeration(elem1);
elem2 = (Element)enum2.nextElement();
if elem2.getAttribute(id) = Num
    while (elem2.getTagname()! = null&&
        elem2.getTagname().toString().unequal
        (“产品”))
    {
        //遍历部件xml文档中的第“Num”
        棵产品子树
    }
}
```

主要通过遍历部件XML文档树直到第“Num”棵“产品”子树,程序通过java.ms.xml.om.Element

(下转第530页)

(上接第 525 页)

类的 `getAttribute(id)` 方法获得标记“<产品>”的“id”属性值,程序把该“id”同从 DBMS 中检索到的零件编号“Num”相比较,若相等,则进一步对产品子树进行遍历.或通过 `java.ms.xml.om.Element` 类的 `getText()` 方法返回“<产品>”标记的所包含的全部非标记信息,即所查找的汽车零件的详细信息.

4 小 结

传统的数据库技术只有与飞速发展的网络技术相适应,才能得以进一步的发展和应用.XML 技

术特别适合于半结构化数据资源的组织和发布.XML 中的文档类型定义 DTD,使人们可以根据实际需要构造所需的标记,不仅方便用户开发基于 Internet 的应用系统,而且结合 Java 技术,可方便地构建一个与用户平台无关的统一而简单的交流方式.本系统基于 JSP 技术,把 XML 技术和 DBMS 技术结合起来,充分发挥传统关系型数据库在信息检索方面的优越性,实现了信息的描述、信息的内容和结构与系统的查询真正地分开.提高了整个系统的网上查询的效率.

参考文献:

- [1] CHRISTOPHER A. Linux Web 编程[M].北京:电子工业出版社,1999.
- [2] HORSTMAN S. Java 2 核心技术[M].北京:机械工业出版社,1999.
- [3] 黄理.用 JSP 轻松开发 Web 网站[M].北京:希望电子出版社,2000.
- [4] HAROLD Rusty 著.XML 实用大全[M].北京:中国水利电力出版社,2000.
- [5] 怀石工作室编著.XML 完全手册[M].北京:中国水利出版社,2000.
- [6] 武苍林著.JDBC 在数据库中的应用[J].计算机应用,1998,18(10):35-36.

(责任编辑:朱明)