

# 基于 DOM 的 XML 解析与应用

蔚晓娟,冉 静,李爱华,尹治本

(西南交通大学 信息科学与技术学院,四川 成都 610031)

**摘 要:**随着 XML 的广泛应用,大量的信息都将通过 XML 文档来进行存储、交换,因此熟悉和掌握一种解析 XML 文档的方法是很重要的。文中着重介绍基于 DOM 的 XML 解析的原理、特点、实现及在数据存储方面的应用,并以业务查询组件为例,详细描述了 XML 建模、解析及与 JAVA 反射机制相结合实现组件的可扩展性,增强了组件的灵活性和扩展性,降低组件的维护成本,突出了 XML 存储数据的优点,对推广应用 XML 和 DOM 技术有一定的实用意义。

**关键词:**文档对象模型;XML 解析;Java 反射机制

**中图分类号:**TP312

**文献标识码:**A

**文章编号:**1673-629X(2007)04-0086-03

## XML Parse and Application Based on DOM

WEI Xiao-juan, RAN Jing, LI Ai-hua, YIN Zhi-ben

(School of Information Science & Technology, Southwest Jiaotong University, Chengdu 610031, China)

**Abstract:** With the widespread use of XML, a lot of information would be stored and exchanged by XML file, so it is very important to familiar and master one kind of XML parse method. Highlights the principle, characteristics and application in the data storage. Taking the bussiness inquiry module as the example, describes detailedly XML modeling, parse and extensibility unifies with the Java reflection mechanism, strengthens flexibility and extensibility of the module, reduces the maintenance cost of the module, highlights the XML merit stored datum. It is practically significant to promote the use of XML and DOM technology.

**Key words:** DOM; XML parse; Java reflection

### 1 XML 技术

XML 是 eXtensible Markup Language(可扩展的置标语言)的缩写,是 W3C 组织于 1998 年 2 月发布的标准。和 HTML 不一样,XML 主要是描述数据本身而不是描述数据的表现格式,所以 XML 文档已经将数据和文档的界限缩小了,可以将 XML 文档看作数据,也可将数据看作 XML 文档。XML 文档很适合于表示存储于数据库中的数据,而一些非结构化文档用 XML 文档存储后,常常能提供更大程度上的信息共享。

XML 作为全球通用的半结构化的数据模型,它的应用领域已广泛涉及出版、电子商务、企业协作以及 Web 服务等方面。由于 XML 具有跨平台、简单易用等特性,在很短的时间内就获得了广泛认同,其应用领域不断地得到拓展,已成为一种被大量使用的通用数据格式。

### 2 文档对象模型 DOM

#### 2.1 DOM 介绍

DOM(Document Object Module)是 W3C 制定的一种独立语言和平台的标准,它提供一个可以通用于各种程序语言、操作系统和应用程序的接口,它定义了文档的逻辑结构以及存取和维护文档的方法。利用 DOM 中的对象,可以对 XML 文档进行读取、搜索、修改、添加和删除等操作。DOM 是对 XML 数据的描述体系,它用树型结构的文档来保存 XML 数据,它包括了对 XML 的解析和处理 XML 数据的 API<sup>[1]</sup>。

#### 2.2 DOM 文档结构

实际上,DOM 就是以树为基础处理 XML 及 HTML 文件的一套 API。标准 W3C 中 DOM 模型有几种类型的节点:Document, Root, Text, Element 以及 Attribute。当使用 DOM 对 XML 文本文件进行操作时,它首先解析文件,将文档中的元素、属性、注释、处理指令都看作节点(Node),然后在内存中以节点树的形式创建 XML 的文件表示<sup>[2]</sup>。DOM 树,除了 Root 或 Document 节点以外的每个节点都有一个父节点。Attribute 节点总是和一个 Element 节点相关联,而用来

收稿日期:2006-06-22

作者简介:蔚晓娟(1982-),女,山西文水人,硕士研究生,研究方向为软件工程;尹治本,博导,教授,研究方向为软件工程理论与技术、网络信息系统分析与设计、算法与计算理论。

封装元素的起始标记和结束标记中的文本是映射到一个 Text 节点的;Text 节点是 Element 节点的一个子节点;Text 节点表示在 XML 文件中的一段文本;Element 节点表示 XML 文件;而 Attribute 节点表示一个元素内部的属性名和值。所以,即使很简单的 XML 文件的表现也可能需要很多种节点类型。

### 3 DOM 解析 XML 文档

#### 3.1 DOM 工作原理

使用 DOM 对 XML 文件进行操作时,首先要解析文件,将文件分为独立的元素、属性和注释等,然后以节点树的形式在内存中对 XML 文件进行表示,就可以通过节点树访问文档的内容,并可以根据需要修改文档,这就是 DOM 的工作原理<sup>[3]</sup>。

#### 3.2 DOM 解析的实现

基于树结构的

XML 解析技术是将结构完整的 XML 文档定义为一棵树,如图 1 所示。XML 文档的组织结构是层层嵌套而完成的,每个 XML 文档都有一个根节点,后跟一个或者一些元素,根节点代表文档本身,元素可以看作根节点的孩子或者是树的分支,孩子元素还可以有孩子元素,从而构成了整个 XML 文档树。在解析 XML 文档树时,处理器从文档内的第一个元素——根元素开始解析,然后解析它面对的每个分枝形成的路径,最终解析整个 XML 文件,将该 XML 文档各元素组成相对应的树形结构,然后向应用程序发送。树是广泛应用的一种数据结构,将 XML 文档解析成树结构以后,许多成熟的算法都可以用来遍历、搜索、编辑 XML 文档树<sup>[4]</sup>。

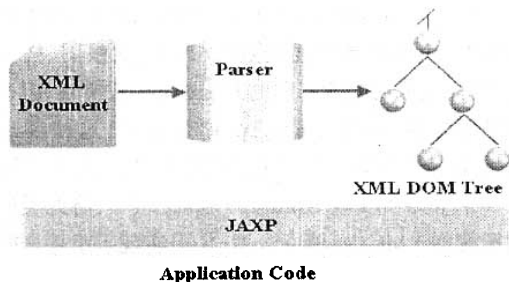


图 1 DOM 解析的实现

DOM 的优点是方便好用,整个文档树在内存中,便于操作;支持删除、修改、重新排列等多种操作功能,因此它被广泛应用。但是,它在使用上却常常会受到

系统资源的限制,尤其是内存。因为它将整个文档调入内存,包括无用的节点,操作浪费时间和空间<sup>[5]</sup>。

### 4 DOM 解析 XML 在信息存储上的应用

XML 是用来定义文档标记语言的框架,主要用来存储和转换数据信息,XML 存储数据时忽略数据类型,使得 XML 非常灵活。XML 存储业务数据时,通过解析 XML 得到业务数据,并对业务数据按一定的接口规则进行处理,如果在规则一定的情况下,业务数据发生变化时,用户可以修改 XML 或修改 XML 解析程序来实现业务与规则的一致,而不需修改接口程序,这使得用户非常方便。如图 2 所示。

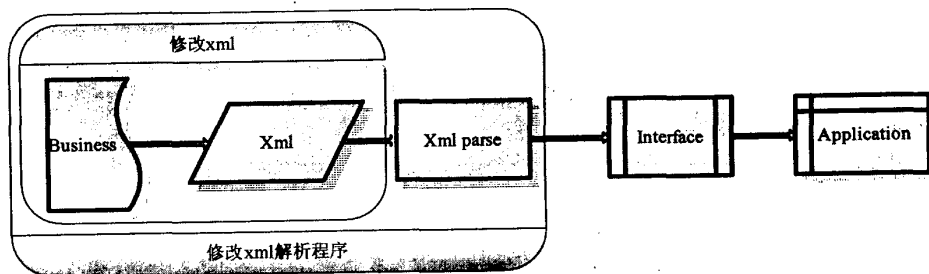


图 2 业务数据变化所引起的变化

业务变化所引起的变化,绝大部分只需要配置 XML 即可完成。对于业务变化,XML 的结构已经不能描述,则需要重新设计 XML 的结构,即 XML 建模,相应的 XML 解析程序需要重写,即便如此,应用程序仍然可以不变。当然有时需要新增一些控件并配置到 XML 中,但原来的应用程序不作任何修改。下面以业务查询组件为例来说明。

#### 4.1 业务查询 XML 建模

XML 建模,即创建描述 XML 文档结构的模型,常用的定义语言有 dtd 和 schema,文中选用 schema 进行 XML 建模,schema 可视化,以便于清晰理解模型树的结构。

XML 建模原则之一,将可分离的信息尽量分离:

- (1) 描述数据库表及表字段的信息与查询所需信息分离。
- (2) 查询条件配置控件独立出来,查询信息中通过名称关联。
- (3) 查询条件配置所需要的数据源信息独立出来。
- (4) 查询结果展现控件独立出来。

按照以上的设计理念,XML 的建模方案,将业务查询 XML 的结构分 5 个主要部分,业务查询 XML 结构如图 3 所示,Tables 存储描述数据库表的信息,包括表名、字段名称、字段类型、表关联(即外表)等信息,属于数据库表固有的信息,与查询无关;ConditionField-

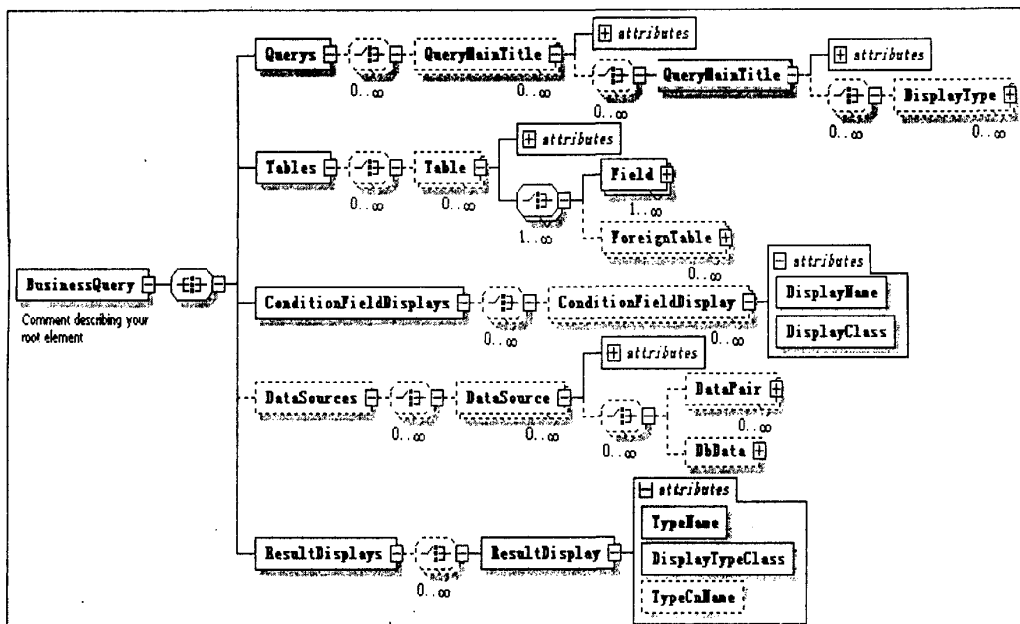


图 3 业务查询 XML 结构模型

Displays 存储描述查询条件配置的控件信息,如选择时间控件、数值输入控件等;DataSources 存储查询条件配置时所需要的供选择的信息,主要分为两类,直接配置到 XML 中的数据对(DataPairs)和数据库查询得到的数据;ResultDisplays 存储描述查询结果展现的控件信息,如常用的表格、曲线图、柱状图;Querys 存储描述查询的信息,查询按主题进行划分,每个查询主题包括多个查询结果展现方式,多个表关联查询,当然也可以是一个表,此节点所需要的条件配置控件、条件配置数据源、查询结果展现控件信息通过名称与其它各节点的信息关联。

#### 4.2 XML 解析

将 XML 模型的结构固化到 Java 的类中,即节点与 Java 类相对应,并增加查找子节点的方法。原因是 XML 结构复杂,如果采用有些文章中介绍的 Hashtable 在内存中存储 XML 的信息,会造成程序难读、难维护,而且会形成 XML 结构与应用程序的强耦合关系,对组件的通用信息有很大的影响。

部分节点增加到父节点对象的引用,如对于字段对象,可以得到该字段属于哪个表对象。根节点 BusinessQuery 增加通过名称查询结果展现、数据源、查询配置等节点的信息。

#### 4.3 Java 反射机制与 XML 结合实现组件的可扩展性

反射(Reflection)是 Java 被视为动态(或准动态)语言的一个关键性质。这个机制允许程序在运行时通过 Reflection APIs 取得任何一个已知名称的 class 的内部信息,包括其 modifiers(诸如 public, static 等等)、su-

perclass(例如 Object)、实现之 interfaces(例如 Cloneable),也包括 fields 和 methods 的所有信息,并可于运行时改变 fields 内容或唤起 methods。要实现组合的可扩展性需要:

- (1)定义 interface,即确定应用程序所需要接口。
- (2)相应的类实现 interface。主要包括 XML 解析类、查询结果展现控件类、条件配置控件类。
- (3)针对接口编程(见表 1)。应用程序通过 interface 调用相应的方法,而不是具体的类。

表 1 应用 Java 反射的代码

```
Class c = Class.forName(className);
SqlResultDisplay sqlPanel = (SqlResultDisplay) c.newInstance();
sqlPanel.setParams(params); //此句是伪代码
JPanel panel = sqlPanel.getPanel();
```

以查询结果展现为例,说明如何实现可扩展性。

ClassName 不用写死到程序中,而是配置到 XML 文档中,显示结果的展现方式(表格、曲线图、柱状图等)可完全采用配置不同的 ClassName 来实现,但这个类必须实现 SqlResultDisplay 这个接口,这就实现了 XML 的可配置性。如果需要新的展现方式,只需要新写一个 Java 类实现 SqlResultDisplay 接口,将此类的类名配置到 XML 中,即可合用,从而实现了组件良好的可扩展性。如图 4 所示。

## 5 结束语

DOM 是一种在浏览器内编程的模型,也是 XML

(下转第 139 页)

### 3 组件使用

建一个部门表 department, 具体字段如下:

department: dept\_id dept\_name remark

以这两个表来介绍该组件的使用流程:

(1) 搭建应用环境。

由于本组件是基于 Struts 和 Hibernate 框架的, 因此首先得下载 Struts 包及其相关配置文件, 并根据其配置规则创建一个 WEB 系统框架。然后下载 Hibernate 包, 创建 Hibernate 的配置文件 hibernate. cfg. xml 和 hibernate. properties, 因为 SessionFactory 通常只需建立一次, 所以可以利用 Struts 的 plugin 机制在系统初始化时建立它。

(2) 根据 Hibernate 的表映射规则, 编写 Department. hbm. xml, 然后建立对应的 Java 映射类 Department。

(3) 在 hibernate. cfg. xml 中建立数据库连接和 SessionFactory, 并申明上面的两个 XML 映射文件。

(4) 对于单张表 Department 进行报表生成列表形式, 如图 2 所示。

### 4 结论

根据面向对象编程思想和设计方法, 把企业报表抽象成一个统一的类, 而它又由 HTML 控件类所组成, 通过 JavaBean 与 js 文件的配合使用, 在 Struts 和 Hibernate 框架中实现了一个可以自动生成的通用企

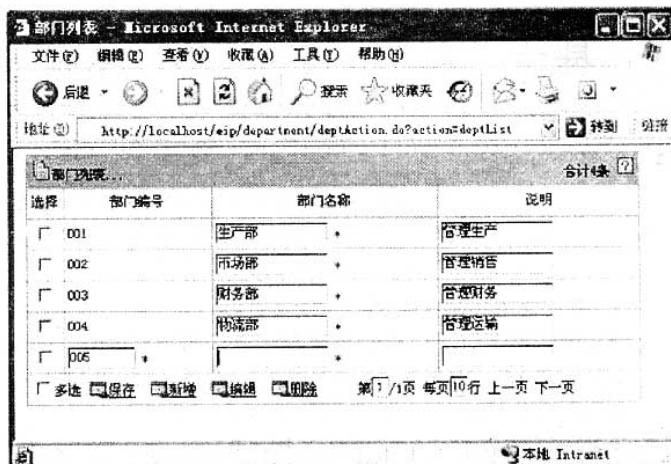


图 2 生成部门列表

业报表, 使得开发者在开发企业报表时, 只需申明需要操作的表的 Hibernate 对象类、页面显示字段、页面显示类型, 大大减小了工作量。

#### 参考文献:

- [1] 孙卫琴. 基于 MVC 的 JavaWeb 设计与开发[M]. 北京: 电子工业出版社, 2004.
- [2] 孙卫琴. 精通 HIBERNATE Java 对象持久化技术详解[M]. 北京: 电子工业出版社, 2005.
- [3] 方 巍, 孙 涌, 张书奎. 整合 Struts 和 Hibernate 的 Web 系统应用[J]. 计算机与现代化, 2005(12): 39-41.
- [4] 高 昂, 卫文学. 基于 Hibernate 与 Struts 框架的数据持久化应用研究[J]. 计算机应用, 2005, 25(12): 2817-2820.
- [5] 阎 宏. Java 与模式[M]. 北京: 电子工业出版社, 2002.

(上接第 88 页)

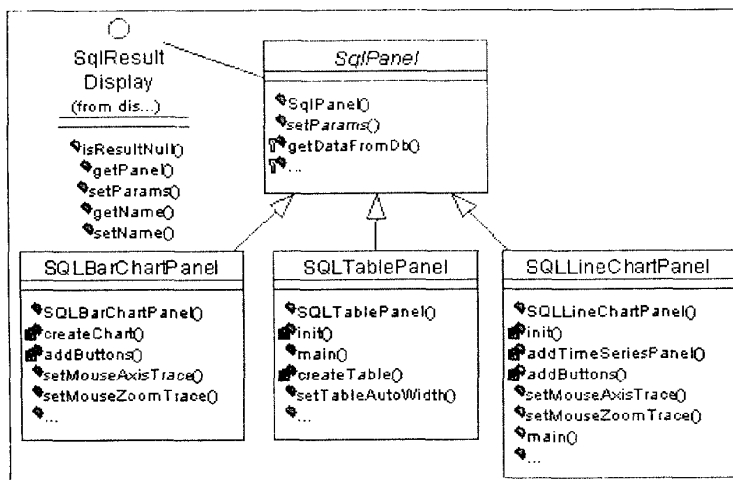


图 4 查询结果展现涉及的接口和类

的主要接口, 是基于树的 API, 它在内存中创建了文档树结构, 通过遍历这颗树来解析到 XML 中存储的数

据。文中着重介绍了 XML 解析在数据存储方面的应用, 对推广应用 XML 和 DOM 技术有一定的实用意义。

#### 参考文献:

- [1] 李勇军, 冀汶莉, 马光思. 用 DOM 解析 XML 文档[J]. 计算机应用, 2001(8): 103-105.
- [2] 邱 俊. 如何有效利用 DOM 处理 XML 文档[J]. 重庆工学院学报, 2002(4): 38-41.
- [3] 朱前飞, 高 芒. XML 解析技术研究[J]. 电脑开发与应用, 2004(11): 26-28.
- [4] 逮 鹏. XML 文档解析技术及其应用[J]. 郑州大学学报, 2002(12): 17-19.
- [5] 孙 霞, 程宏斌. 基于 JAVA 的 DOM 解析技术[J]. 计算机时代, 2004(7): 1-3.