

深入解析dom4j

——以面向对象的视角来看

作者：赵飞 更新日期：2018年11月1日 基于国科大面向对象程序设计课程



Part1: dom4j的功能与使用

1. dom4j的简介

- 本手册分析版本：[v2.1.1](#)
- dom4j维护官网地址：[dom4j官网](#)，github仓库地址：[github仓库](#)
- **先看一个dom4j的官方释义**：Dom4j is an easy to use, open source library for working with XML, XPath and XSLT on the Java platform using the Java Collections Framework and with full support for DOM, SAX and JAXP.
- 是的，dom4j是一个**开源的**、基于**Java语言**的**类库**。它专用于在Java平台上处理XML, XPath,XSLT等类型的文件（本手册仅关注于该类库在处理解析**XML**文件方面的作用）。这里不再赘述dom4j的各种优越性，我们只需要知道：dom4j是一个使用简单，执行高效的XML解析包，就连Sun公司的JAXM也在使用它解析XML文件。现今，dom4j已经成为jar包的一个标配。

2. dom4j的XML解析功能

- dom4j，顾名思义，它就是支持对XML文件进行**DOM操作**的一个类库。那么什么是DOM操作呢？事实上，DOM操作就在我们每个人的日常生活中。你每天浏览网页（静态网页除外）其实就是在执行DOM操作。
- 首先，我们先来看看XML文件的结构。众所周知，浏览器中的网页大多数都是由HTML语言（XML语法和HTML语法非常类似）描述的，例如下面就是我喜欢的一个wap网页游戏的部分首页**HTML**源代码：

```
1  <!--HTML文件结构-->
2  <?xml version="1.0" encoding="utf-8"?>
3  <!DOCTYPE html PUBLIC "-//WAPFORUM//DTD XHTML Mobile 1.0//EN"
   "http://www.wapforum.org/DTD/xhtml-mobile10.dtd">
4  <html>
5  <head>
6      <title>大话水浒</title>
7      <script src="http://203.195.129.90/html/js/wml2html.js"></script>
8      <link href="http://203.195.129.90/html/css/gamestyle.css"
   type="text/css" rel="stylesheet"/>
9  </head>
```

```

10 <body>
11 <div id="main_body">
12 <p>
13 
14 <br />
15 帐号:
16 <br />
17 <input name="username" type="text" value=""/>
18 </div>
19 </body>
20 </html>

```

再看看dom4j自带的一个示例XML文件 `\dom4j-master\xml\test\journal.xml` :

```

1 <?xml version="1.0"?>
2
3 <series xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
4       xsi:noNamespaceSchemaLocation="journal.xsd">
5
6 <entry>
7 <date>May 19, 2002</date>
8 <text>
9 Sample text. We should support some markup here.
10 </text>
11 </entry>
12
13 <entry>
14 <date>May 19, 2002 10:27</date>
15 <text>
16 Second batch.
17 </text>
18 </entry>
19
20 </series>

```

可以看到，XML，HTML它们都由层次化的标签（tag）+内容（content）组成。例如在journal.xml中文档开头是 `<?xml version="1.0"?>`，意在声明这是一个XML格式的文件。紧接着就是嵌套的多个元素（元素=tag+content）。在一个元素中，tag名描述了其中内容的含义、类型。例如其中的 `<date>May 19, 2002 10:27</date>`，标签名就是date（以`<`包装为开始标签，以`</>`包装为结束标签），它表示这个元素的内容19,2002 10:27是一个日期（date）。

3. dom4j的大概组成

4. dom4j的使用方法

Part2: dom4j的核心类、对象、运作原理

Part3: dom4j的高级设计意图

最近实验多比较忙，敬请期待后续良心更新...

