

XML 与 HTML 整合

陈银凤

(内蒙古财经学院计算机信息管理学院, 呼和浩特 010051)

摘 要: XML 与 HTML 都是重要的网络标识语言,但各自具有优缺点,XML 侧重于数据的描述,HTML 侧重于数据的显示,若将两者整合则可以取长补短。主要介绍如何使用数据岛技术来实现 XML 与 HTML 的整合。

关键词: XML; HTML; 整合; 数据岛

0 引言

XML 擅长于数据描述,HTML 擅长于数据显示,这两种标识语言都在目前的 Web 中占有重要地位。但 XML 的出现不会取代 HTML,而是相辅相成,互为补充。因此理想的方案是把 XML 和 HTML 两种技术直接相结合,优势互补,用户可以利用 XML 来描述数据和存储数据,利用 HTML 来显示数据。

1 相关概念

HTML(Hypertext Markup Language),超文本标识语言,是构建网页的描述语言,提供了精简而有力的文件定义,用户可以利用它来设计出多姿多彩的超媒体文件,通过 HTTP 协议,使得 HTML 文件可以在万维网上进行跨平台的文件交换。

XML(eXtended Markup Language),可延伸标记语言或可扩展标记语言,与 HTML 一样,都是 SGML(Standard Generalized Markup Language,标准通用标记语言)。XML 提供了一套跨平台、跨网络、跨程序语言的数据描述方式,各行各业的人都可以针对自己的需求定义所要的文件格式,并且以 XML 文件当作数据交换的标准格式。

2 XML 文件的特性

XML 与先期的 HTML,虽然同为 SGML 的一种应用,但是由于 XML 的突出优点,而且它能克服 HTML

的许多缺点,例如 HTML 标记语言仅提供一种显示信息的方式;标签是预先定义好的,不可扩展;不具备任何高层次标记语言能力,不能理解文档概念,不具备结构化,不利于信息存储等。人们开始研究用 XML 文件来存储组织管理 Internet 上庞大的信息资源。而 XML 具有良好的扩展性;XML 是面向数据而非面向显示的,其内容与形式相互分离、内容易识别、修改方便、数据显示灵活;XML 是自我描述语言,具有严格的语法要求,这使得 XML 文件清晰可读。数据的树形组织形式很好地体现了信息的层次性及复杂的树状关系和继承关系;XML 支持世界上大多数文字,可以标注各种文字、图像及二进制文件,可以轻松实现跨平台、跨国界的信息交换。

3 XML Data Islands

在 HTML 网页中可以利用数据链路的技术,将外部数据整合到 HTML 文件中,称这些嵌入的数据为 Data Islands(数据岛)。

XML Data Islands(XML 数据岛)是指存在于 HTML 页面中的 XML 数据,就是使用<XML>标记嵌入 XML 数据,在 HTML 文档中形成的一个 XML 数据岛。它将 HTML 和 XML 技术直接结合。数据岛技术不是 W3C 的推荐标准,它是微软的技术,在 Internet Explorer 5.0 版本以上的浏览器中才可以使用,通过该浏览器,非正式的<XML>标记可用于创建一个 XML 数据岛。

收稿日期:2011-05-17 修稿日期:2011-06-17

作者简介:陈银凤(1978-),女,内蒙古临河人,讲师,研究方向为计算机网络、电子商务、数据库

HTML 有两种方式引入 XML 数据岛,一种为嵌入式,另一种为链入式。如果是要直接加入 XML 文件,那就在<XML>和</XML>两个标记之间,放入整个 XML 文件,并包含 XML 的版本声明。通常在 XML 元素中加入 ID 属性,以方便以后在 HTML 中读取 XML 数据岛的内容。这种方式即为嵌入式,如例 1 所示。如果是要参照到一个外部的 XML 文件,那就必须利用 XML 元素的 SRC 属性。SRC 属性的属性值可以是一个本地的 XML 文件名称,或是一个 URL 位置,以指到一个外部的服务器,这种方式即为链入式,如例 2 所示。

例 1:

```
<html>
<body>
  <xml id="XMLID">
    <?xml version="1.0" encoding="gb2312"?>
      <book>
        <bookname>XML 技术</bookname>
        <author>chenYF</author>
      </book>
    </xml>
    <!--使用数据岛数据的 HTML 代码-->
  </body>
</html>
```

例 2:

```
<html>
<body>
  <xml id="XMLID" src="book.xml"/>
  <!--使用数据岛数据的 HTML 代码-->
</body>
</html>
```

目前许多商务网站都尽可能把商务规则、原始数据和表现形式,即执行功能的代码、存放数据的数据库、显示数据的页面单独地划分出来,做成相互独立的服务。因此,当前的做法把从数据库提取出来的数据存放在 XML 文件中,使用 XSL 来显示数据内容;或者把从数据库提取出来的数据以数据岛的形式引入到 HTML 文档中,既能实现数据和显示分离,又能充分利用 HTML 的显示优势。

4 Data Source Object

Data Source Object(数据源对象,英文简称 DSO)

对象主要是提供数据给 HTML 网页,程序设计者可以将 DSO 对象视为简化的 ADO Recordset 对象,可以使用 ADO 对象的方法来处理元素的显示。

为了能够处理内嵌 XML 代码的 HTML 页面,Internet Explorer 4.0 及其以上的版本引入了 DSO 技术。XML DSO 是一个 Microsoft ActiveX 控件,嵌入到 Microsoft Internet Explorer 中,利用嵌入的控件,DSO 有助于格式化数据并将数据显示在 HTML 页面上。

DSO 对象保存了 XML 文件的数据,并提供了访问数据的方法。DSO 将 XML 文件看作数据库文件,被绑定的 XML 一般具有三层结构:

①第一层是根元素,相当于数据表;

②第二层为根元素的子元素,第二层中每个元素具有完全相同的子元素,这一层相当于数据表中的记录;

③第三层元素相当于字段。

DSO 将 XML 数据存储成记录集,因此较适用于显示那些具有对称性数据记录集结构的 XML 文件。数据绑定实际上是将 XML 的字段与 HTML 的指定标记进行绑定。绑定的方式按照 HTML 元素的性质可分为两种:单值对象绑定和表格对象绑定。单值绑定只能显示一条记录的数据;表格绑定可以显示所有记录的数据。

(1)单值对象绑定

XML 文件的第二层相当于数据表中的记录,单值对象绑定只能显示一条记录的数据。单值对象的数据绑定需要在 HTML 元素中设置 Datasrc 属性和 Datafld 属性。前者是区分数据岛对象的唯一标识,并且必须加上前缀 #,它相当于数据表对象;后者则标识了数据岛中的元素对象,对应于第三层元素,相当于字段。如例 4 对例 3 的调用。

例 3: 建立 bookdata.xml, 一个存储了 2 本书的 XML 文件。

```
<?xml version="1.0" encoding="gb2312"?>
<bookdata>
  <book>
    <bookname>数据库原理</bookname>
    <author>王珊</author>
  </book>
  <book>
    <bookname>计算机网络</bookname>
    <author>吴功宜</author>
  </book>
```

<bookdata>

例 4:

```
<html>
<head><title>单值对象绑定</title></head>
<body>
<xml id="dos" src="bookdata.xml"></xml>
<center>
<h1>下面的数据是从 XML 提取出来的</h1>
<p><span datasrc="#dos" datafld="书名"></span></p>
<p><span datasrc="#dos" datafld="书名"></span></p>
</center>
</body>
</html>
```

(2) 表格对象绑定

表格对象绑定主要是与 HTML 中的<TABLE>标记绑定。表格对象绑定在显示的时候,可以通过数据源对象 DSO 遍历 XML 里面的所有记录,再借助 DSO 等 Active 控件能够更方便、有效地处理表格数据。在进行数据绑定时,整个表格对应一个记录集,表格中的每一行对应一条记录,表格中的每个单元格对应一个字段。因此需要在<TABLE>标记设置 datasrc 属性值为 XML 对象的 id 值,在单元格内设置 datafld 属性,指定与 XML 绑定的字段,例如:

例 5:

```
<html>
<head><title>表格对象绑定</title></head>
<body>
<xml id="dos" src="bookdata.xml"></xml>
<center>
<table datasrc="#dos" border="1" width="80%">
<thead>
```

```
<tr>
<th>书名</th>
<th>编者</th>
</tr>
</thead>
<tr>
<td><span datafld="书名"></span></td>
<td><span datafld="书名"></span></td>
</tr>
</table>
</center>
</body>
</html>
```

5 结 语

HTML 提供了丰富的显示页面,但 HTML 没有将数据的存储和数据的显示分离,不仅使数据难以重用,而且使系统的可扩展性差。而 XML 文件仅仅保存了数据,没有描述数据的显示方式,因此可以将 XML 与 HTML 进行整合,即用 XML 来存储、检索结构化和非结构的 XML 数据,并利用 HTML 方便地在 Web 中显示出来。目前,XML 与 HTML 在电子商务领域及 Web 数据库中已经有大量实际应用。

参考文献

- [1]林锦雀. 最新 XML 入门与应用[M]. 北京:中国铁道出版社,2001.11
- [2]焦阳. XML 与 HTML 的整合技术[J]. 现代图书情报技术, 2004,115:47~49
- [3]蔡体健. XML 网页设计实用教程[M]. 北京:人民邮电出版社,2009.3

Integration of XML and HTML

CHEN Yin-feng

(Department of Computer Information and Management, Inner Mongolia Finance and Economics College, Hohhot 010051)

Abstract: Both XML and HTML are the important network identifier languages, but each has its own advantages and disadvantages. XML is focused on data description and HTML is focused on data display. If XML and HTML can be integrated, it can complement each other. Mainly introduces how to use data islands technology to realize the integration of XML and HTML.

Keywords: XML; HTML; Integration; Data Islands

作者: [陈银凤](#)
作者单位: [内蒙古财经学院计算机信息管理学院, 呼和浩特, 010051](#)
刊名: [现代计算机: 下半月版](#)
英文刊名: [Modern Computer](#)
年, 卷(期): 2011(13)

参考文献(3条)

1. [林锦雀](#) [最新XML入门与应用](#) 2001
2. [焦阳](#) [XML与HTML的整合技术](#) 2004
3. [蔡体健](#) [XML网页设计实用教程](#) 2009

本文链接: http://d.wanfangdata.com.cn/Periodical_xdjsj-xby201113014.aspx