

6.6 解析和修改XML

问题

你想读取一个XML文档，对它做一些修改，然后将结果写回XML文档。

解决方案

使用 `xml.etree.ElementTree` 模块可以很容易地处理这些任务。第一步是以通常的方式来解析这个文档。例如，假设你有一个名为 `pred.xml` 的文档，类似下面这样：

```
<?xml version="1.0"?>
<stop>
  <id>14791</id>
  <nm>Clark & Balmoral</nm>
  <sri>
    <rt>22</rt>
    <d>North Bound</d>
    <dd>North Bound</dd>
  </sri>
  <cr>22</cr>
  <pre>
    <pt>5 MIN</pt>
    <fd>Howard</fd>
    <v>1378</v>
    <rn>22</rn>
  </pre>
  <pre>
    <pt>15 MIN</pt>
    <fd>Howard</fd>
    <v>1867</v>
    <rn>22</rn>
  </pre>
</stop>
```

下面是一个利用 `ElementTree` 来读取这个文档并对它做一些修改的例子：

```
>>> from xml.etree.ElementTree import parse, Element
>>> doc = parse('pred.xml')
>>> root = doc.getroot()
>>> root
<Element 'stop' at 0x100770cb0>

>>> # Remove a few elements
>>> root.remove(root.find('sri'))
>>> root.remove(root.find('cr'))
>>> # Insert a new element after <nm>...</nm>
>>> root.getchildren().index(root.find('nm'))
1
>>> e = Element('spam')
>>> e.text = 'This is a test'
>>> root.insert(2, e)

>>> # Write back to a file
>>> doc.write('newpred.xml', xml_declaration=True)
>>>
```

处理结果是一个像下面这样新的XML文件：

```
<?xml version='1.0' encoding='us-ascii'?>
<stop>
  <id>14791</id>
  <nm>Clark & Balmoral</nm>
  <spam>This is a test</spam>
  <pre>
    <pt>5 MIN</pt>
```

```
<fd>Howard</fd>
<v>1378</v>
<rn>22</rn>
</pre>
<pre>
<pt>15 MIN</pt>
<fd>Howard</fd>
<v>1867</v>
<rn>22</rn>
</pre>
</stop>
```

讨论¶

修改一个XML文档结构是很容易的，但是你必须牢记的是所有的修改都是针对父节点元素，将它作为一个列表来处理。例如，如果你删除某个元素，通过调用父节点的 `remove()` 方法从它的直接父节点中删除。如果你插入或增加新的元素，你同样使用父节点元素的 `insert()` 和 `append()` 方法。还能对元素使用索引和切片操作，比如 `element[i]` 或 `element[i:j]`

如果你需要创建新的元素，可以使用本节方案中演示的 `Element` 类。我们在6.5小节已经详细讨论过了。