1. **Python使用SAX解析XML**

**第一部分**

SAX是一种基于事件驱动的API,通常包含三个事件：开始解析标签事件、解析数据事件和结束解析标签事件。

利用SAX解析XML文档牵涉到两个部分:解析器和事件处理器。解析器负责读取XML文档,并向事件处理器发送事件,如元素开始和元素结束事件;而事件处理器则负责对事件作出相应数据进行处理。

SAX解析XML适合的场景：

 1、对大型文件进行处理；

 2、只需要文件的部分内容，或者只需从文件中得到特定信息。

**一、ContentHandler类**

在python中使用sax方式处理xml要先引入xml.sax中的parse函数，还有xml.sax.handler中的ContentHandler。

**1、ContentHandler类方法介绍**

**1.1 characters(content)方法**

调用时机：

从行开始，遇到标签之前，存在字符，content的值为这些字符串。

从一个标签，遇到下一个标签之前， 存在字符，content的值为这些字符串。

从一个标签，遇到行结束符之前，存在字符，content的值为这些字符串。

标签可以是开始标签，也可以是结束标签。

**1.2、startDocument()方法**

文档启动的时候调用。

**1.3、endDocument()方法**

解析器到达文档结尾时调用。

**1.4、startElement(name, attrs)方法**

遇到XML开始标签时调用，name是标签的名字，attrs是标签的属性值字典。

**1.5、endElement(name)方法**

遇到XML结束标签时调用。

1. **解析器**

**2.1 创建解析器**

以下方法创建一个新的解析器对象并返回。

xml.sax.make\_parser( [parser\_list] )

参数说明:

 parser\_list - 可选参数，解析器列表

**2.2 、parser方法**

以下方法创建一个 SAX 解析器并解析xml文档：

xml.sax.parse( xmlfile, contenthandler[, errorhandler])

参数说明:

 xmlfile - xml文件名

 contenthandler - 必须是一个ContentHandler的对象

 errorhandler - 如果指定该参数，errorhandler必须是一个SAX ErrorHandler对象

查看xml.sax.ContentHandler类的内容

>>> import xml.sax

>>> help(xml.sax.ContentHandler)

**第二部分**

**三、Python 解析XML实例**

#encoding=utf-8

import xml.sax

import sys

import xml.sax.handler

#1、自定义一个studentHander，使其继承xml.sax.ContentHandler

class studentHander(xml.sax.ContentHandler):

def \_\_init\_\_(self):

print("解析开始：")

#在开始解析标签事件中应处理的业务逻辑

#其中name表示解析的标签名 ，attrs表示解析标签的属性集合

def startElement(self, name, attrs):

if name=="student":

print("<%s sid=%s>" %(name,attrs["sid"]),end="")

else:

print("<%s>" %(name),end="")

#在结束解析标签事件中应处理的业务逻辑

def endElement(self, name):

print("</%s>" %(name))

#解析数据(文本)事件中执行的业务逻辑

def characters(self, content):

print("%s" %(content),end="")

#2、创建一个新的解析器对象并返回

parser=xml.sax.make\_parser()

#3、指定解析器的ContentHandler对象

parser.setContentHandler(studentHander() )

#4、解析xml文件

parser.parse("Students.xml")

执行结果：

遍历开始：

<students>

<student sid=1>

<id>1</id>

<name>陈丽水</name>

<sex>女</sex>

<address>北京昌平</address>

<age>18</age>

</student>

<student sid=2>

<id>2</id>

<name>方军</name>

<sex>男</sex>

<address>北京海淀</address>

<age>20</age>

</student>

<student sid=3>

<id>3</id>

<name>李华</name>

<sex>女</sex>

<address>北京通州</address>

<age>19</age>

</student>

<student sid=4>

<id>4</id>

<name>淮海</name>

<sex>男</sex>

<address>北京顺义</address>

<age>22</age>

</student>

<student sid=5>

<id>5</id>

<name>成功</name>

<sex>男</sex>

<address>北京海淀</address>

<age>18</age>

</student>

</students>