



python 魔鬼训练营 第9周

DATAGURU专业数据分析社区

【声明】 本视频和幻灯片为炼数成金网络课程的教学资料，所有资料只能在课程内使用，不得在课程以外范围散播，违者将可能被追究法律和经济责任。

课程详情访问炼数成金培训网站

<http://edu.dataguru.cn>

第9课：XML就该这么用

课程内容：

- XML简介
- XML库介绍
- 库使用

本次课内容重点讲解日常工作中使用python如何进行XML的读写操作；
重点学习xx库的使用。

第1节：XML简介

- XML：全称可扩展标记语言；在电子计算机中，标记指**计算机**所能理解的信息符号，通过此种标记，计算机之间可以处理包含各种的信息比如文章等。它可以用来标记数据、定义数据类型，是一种允许用户对自己的标记语言进行定义的源语言。它非常适合**万维网**传输，提供统一的方法来描述和交换独立于**应用程序**或供应商的结构化数据。是Internet环境中跨平台的、依赖于内容的技术，也是当今处理分布式结构信息的有效工具。早在1998年，W3C就发布了XML1.0规范，使用它来简化Internet的文档信息传输。
- 作用：为不同应用提供一种结构化的数据存储和传输的方式。

第2节：XML库介绍

- xml.dom : xml三剑客
- xml.etree : xml三剑客
- xml.sax : xml三剑客
- expat : 号称最快的xml解析器
- libxml2 : 比较基础的xml解析库
- lxml.etree : 是libxml2和libxslt库的一个Python化的绑定
- BeautifulSoup : 底层调用 lxml , htmlib , htmlparser

第3节：xml.dom库使用

- minidom模块
- Document工厂方法
- DOM对象方法
- 创建xml节点
- 创建xml属性
- 追加内容
- 回显xml
- 写入xml文件

第3节：xml.dom库使用

- 读取xml文件
- 获取根节点
- 获取子节点
- 获取节点类型
- 获取节点名称
- 获取节点值
- 获取节点属性

第3节：xml.etree库使用

- 插入节点
- 删除节点
- 设置节点属性
- 查找节点
- 获取节点属性
- 获取节点文本

第3节：xml.sax库使用

- ContentHandler
- startElement
- endElement
- characters
- startdocument
- Enddocument
- parse
- parseString

第9课：XML就该这么用

要点回顾：

- XML库说明
- 重点库的使用

- Dataguru (炼数成金) 是专业数据分析网站 , 提供教育 , 媒体 , 内容 , 社区 , 出版 , 数据分析业务等服务。我们的课程采用新兴的互联网教育形式 , 独创地发展了逆向收费式网络培训课程模式。既继承传统教育重学习氛围 , 重竞争压力的特点 , 同时又发挥互联网的威力打破时空限制 , 把天南地北志同道合的朋友组织在一起交流学习 , 使到原先孤立的学习个体组合成有组织的探索力量。并且把原先动辄成千上万的学习成本 , 直线下降至百元范围 , 造福大众。我们的目标是 : 低成本传播高价值知识 , 构架中国第一的网上知识流转阵地。
- 关于逆向收费式网络的详情 , 请看我们的培训网站 <http://edu.dataguru.cn>

Thanks

FAQ时间