

网络编程

- 网络编程
 - 语言基础
 - socket
 - Python中的套接字
 - 模块urllib使得HTTP请求
 - 解析HTML (Web抓取)
 - web数据
 - XML
 - JavaScript Object Notation
 - 面向服务方法
 - Service Oriented Approach
 - Application Program Interface

语言基础

socket

- TCP链接 / Sockets 在计算机网络中，socket是指基于因特网协议的计算机网络之上的 全双工通信流的一端。
- Common TCP Ports: telnet: 23 ssh: 22 http: 80 https: 443 smtp: 25 dns: 53 ftp: 21

Python中的套接字

Python有模块支持TCP Sockets。

```
import socket
mysock=socket.socket(socket.AF_INET, socket.SOCK_STREAM)
mysock.connect( ('python.org', 80) )
```

当使用sockets与网络资源对话或者 使用网络数据库，必须把数据编码 (通常是 UTF-8)，然后解码

模块urllib使得HTTP请求

```
import urllib.request, urllib.parse, urllib.error
fhand = urllib.request.urlopen('http://www.poly.com.cn/s/1077-3767-17480.html')
for line in fhand:
    print(line.decode().strip())
```

解析HTML (Web抓取)

简单办法- Beautiful Soup分析字符串。

web数据

- 在HTTP协议的支持下，程序之间使用协议交换数据就很自然了。需要在应用进程之间、网络中协商一种表示数据的方法，通常使用两种形式: XML and JSON。

XML

eXtensible Markup Language，主要目的是帮助信息系统共享结构化数据。

- 空格：有无换行或者缩进都无所谓， 缩进仅为了可读。
- XML 术语： Tags 指示元素的开始和结束； Attributes-XML的tag中 Keyword/value 对； Serialize/De-Serialize-使用编程语言把程序中的数据转化成共同格式，这种格式可以存储并在系统间传递。
- XML 框架： XML 文档合法形式的描述。描述了文档结构和内容的限制，通常用于指出系统之间的“协议”-“系统仅接受遵循特定框架的XML”，一个XML文件符合框架的规范，称为“合法的”。

XSD XML 框架(W3C spec) 聚焦于the World Wide Web Consortium (W3C)版本 也经常称为“W3C Schema”，因为认为“Schema”是通用的，更多的时候称为 XSD，因为文件名以 .xsd结尾。

JavaScript Object Notation

JSON使用嵌套的“列表”和“字典”表示数据。

面向服务方法

Service Oriented Approach

许多重要的web应用都使用服务，从其他应用程序中使用服务。Service公布自己的规则，应用程序必须遵循这些规则才能使用服务 (API)。

Application Program Interface

API本身很抽象，通过接口可以控制接口所指对象的行为。使用API功能的软件称为API的实现。通常会根据编程语言来定义API，这种编程语言用来构建应用程序。