Contents

[1预备知识 3](#_Toc500754016)

[1.1 关于本指导手册 3](#_Toc500754017)

[1.2 依赖条件 3](#_Toc500754018)

[2介绍Python 3](#_Toc500754019)

[2.1 什么是Python? 3](#_Toc500754020)

[2.2 体验Python 4](#_Toc500754021)

[2.3 为什么使用Python? 5](#_Toc500754022)

[3 Python 套接字模块 6](#_Toc500754023)

[3.1 基本Python 套接字模块 6](#_Toc500754024)

[3.2 Socket 6](#_Toc500754025)

[3.3 SocketServer 8](#_Toc500754026)

[4 Python套接字编程 9](#_Toc500754027)

[4.1 依赖条件 9](#_Toc500754028)

[4.2创建和销毁套接字 9](#_Toc500754029)

[4.3套接字地址 10](#_Toc500754030)

[4.4服务器套接字 10](#_Toc500754031)

[4.5 客户端套接字 11](#_Toc500754032)

[4.6 流套接字I/O 12](#_Toc500754033)

[4.7 数据报套接字I/O 13](#_Toc500754034)

[4.8 套接字选项 14](#_Toc500754035)

[4.9 异步I/O 14](#_Toc500754036)

[5构建一个Python聊天服务器 16](#_Toc500754037)

[5.1 一个简单的聊天服务器 16](#_Toc500754038)

[5.2 ChatServer类 16](#_Toc500754039)

[5.3 run方法 18](#_Toc500754040)

[5.4 Helper方法 19](#_Toc500754041)

[5.5实例化一个新的ChatServer 20](#_Toc500754042)

[5.6 验证ChatServer 20](#_Toc500754043)

[6 高级网络类 22](#_Toc500754044)

[6.1 网络模块 22](#_Toc500754045)

[6.2 httplib(HTTP 客户端) 22](#_Toc500754046)

[6.2.1 HTTPConnection类 23](#_Toc500754047)

[6.2.2 httplib.HTTPResponse 24](#_Toc500754048)

[6.2.3 补充说明 25](#_Toc500754049)

[6.3 smptlib (SMTP客户端) 25](#_Toc500754050)

[6.4 poplib(POP3 客户端) 27](#_Toc500754051)

[6.4.1 poplib类 28](#_Toc500754052)

[7总结 29](#_Toc500754053)

Python套接字编程

摘要：使用Python基本套接字功能开发网络应用

# 1预备知识

## 1.1 关于本指导手册

Python是一种非常流行的面向对象的脚本语言，具有语法简洁的特点，并且拥有非常广泛的开发者基础。它是通用目的语言，能够在各种设定中使用。它在刚开始编程的初学者也非常受欢迎，就是20世纪70年代的BASIC语言那样。

本教程的侧重点是Python语言在网络编程方面的应用。 介绍了Python的基本套接字，除此之外，还提供异步套接字的一些Python的其他类。 我还详细介绍了Python的应用层协议类，展示了如何构建Web客户端，邮件服务器和客户端等等。

我还演示了一个简单的聊天服务器来说明Python套接字在套接字应用方面的强大。

在阅读本文前，你应该对标准BSD套接字APIs有一定的理解，且有GNU/Linux环境下工作的经验。熟悉面向对象编程也是非常有益的。

## 1.2 依赖条件

本教程和演示中的示例依赖于Python的2.4版本。 您可以从Python官方网站下载此版本（[请参阅参考资料中的链接](https://www.ibm.com/developerworks/linux/tutorials/l-pysocks/#resources)）。 要构建Python解释器，需要GNU C编译器（gcc）和configure / make工具（它们是任何一种标准GNU / Linux发行版的一部分）。

# 2介绍Python

首先，让我们体验一下Python。

## 2.1 什么是Python?

Python是一个通用的面向对象的脚本语言，可以应用于各种各样的问题。 它是在20世纪90年代初期在阿姆斯特丹的CWI创建的，其后在Python软件基金会下继续发展到今天。

Python有着不可思议的移植性，几乎可以在所有的操作系统上发现它的身影。

Python是解释型语言，且很容易扩展。你可以通过添加模块扩展Python，Python模块可以包含函数，变量，或者通过C/C++编译的函数类型。

您还可以轻松地将Python嵌入到C或C ++程序中，使您可以使用脚本功能扩展应用程序。

Python最有用的方面之一就是其拥有大量的扩展模块。这些模块不仅提供了诸如字符串或列表处理的标准功能，还提供了用于视频和图像处理，音频处理以及网络的应用层模块。

## 2.2 体验Python

现在让我来告诉你Python是什么吧！

作为一种解释型语言，使用Python轻松实现想法并快速建立原型软件非常简单。 Python程序可以整体解释，或一行一行解释。

您可以通过先运行Python来测试以下Python代码片段，然后逐个输入一行。 Python被调用后，出现一个提示符（>>>），允许您输入命令。 请注意，缩进在Python中很重要，因此一行中的前面的空格不能被忽略：

例1．尝试一些Python例子

# 打开一个文件, 逐行读取,然后打印出来

for line in open('file.txt'):

print line

# 创建一个新文件，写入内容

file = open("test.txt", "w")

file.write("test line\n")

file.close()

# 创建字典，包含名字和年龄，

family = {'Megan': 13, 'Elise': 8, 'Marc': 6}

# 读取8

family['Elise']

# 移除一个键值对

del family['Elise']

# 创建一个列表和函数，函数对列表中的每个元素求平方。

# 将函数与列表中的元素建立键值对关系。

arr = [1, 2, 3, 4, 5]

def double(x): return x\*x

map(double, arr)

# 通过继承创建一个类，并实例化，且调用它的方法

class Simple:

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

self.name = name

def hello(self):

print self.name+" says hi."

class Simple2(Simple):

def goodbye(self):

print self.name+" says goodbye."

me = Simple2("Tim")

me.hello()

me.goodbye()

## 2.3 为什么使用Python?

学习和使用Python的头号原因是它的受欢迎程度。其用户群的规模以及使用Python构建的应用程序数量不断增加，这是一项有价值的投资。

您可以在几个开发领域找到Python，例如，构建系统实用程序，程序集成，Internet应用程序和快速搭建原型程序等方面。

与其他脚本语言相比，Python也有一些优势。 它的语法简单，概念清晰，易于学习。 使用复杂的数据结构（如列表，字典和元组）时，Python也更容易且更具描述性。 Python也可以扩展语言，反过来也可以被语言扩展。

我发现Python的语法比Perl更具可读性和可维护性，但不如Ruby。Python相对于Ruby的优势在于可用的大量库和模块。 使用这些，您可以使用少量自定义代码来构建功能丰富的程序。

Python使用缩进来标识块范围可能相当烦人，但其简单性往往弥补了这个小缺陷。

现在，我们来深入Python中的套接字编程吧。

# 3 Python 套接字模块

## 3.1 基本Python 套接字模块

Python提供了两个基本的套接字模块。 首先，Socket提供了标准的BSD套接字API。 第2个，SocketServer提供了一个以服务器为中心的类，它简化了网络服务器的开发。 它以异步方式执行此操作，您可以在其中提供插件类来执行服务器的特定于应用程序的作业。 表1列出了本节涵盖的类和模块。

表1. Python类和模块

|  |  |
| --- | --- |
| 类/模块 | 描述 |
| Socket | 低级网络接口（根据BSD API） |
| SocketServer | 提供简化网络服务器开发的类 |

让我们来看看这些模块，了解他们可以为您做些什么。

## 3.2 Socket

Socket模块提供了UNIX程序员最熟悉的基本网络服务（也称为BSD API）。 这个模块拥有构建套接字服务器和客户端所需的一切。

这个API和标准C API的区别在于它的面向对象。 在C中，从套接字调用中检索套接字描述符，然后将其用作BSD API函数的参数。 在Python中，套接字方法返回套接字方法可应用的套接字对象。 表2提供了一些类方法，表3显示了实例方法的一个子集。

表2. Socket模块的类方法

|  |  |
| --- | --- |
| 类方法 | 描述 |
| Socket | 底层网络接口（BSD API） |
| socket.socket(family, type) | 创建并返回一个新的套接字对象 |
| socket.getfqdn(name) | 将字符串IP地址转换为完全限定的域名 |
| socket.gethostbyname(hostname) | 根据主机名得到IP地址字符串 |
| socket.fromfd(fd, family, type) | 根据现有的文件描述符创建一个套接字对象 |

表3. Socket模块的实例化方法

|  |  |
| --- | --- |
| 实例化方法 | 描述 |
| sock.bind((adrs, port) ) | 将套接字绑定到地址和端口 |
| sock.accept() | 返回一个客户端套接字（带有对等地址信息） |
| sock.listen(backlog) | 将套接字置于监听状态，能够挂起未完成的连接请求 |
| sock.connect((adrs, port) ) | 将套接字连接到定义的主机和端口 |
| sock.recv( buflen[, flags] ) | 从套接字接收数据，最多为buflen字节 |
| sock.recvfrom( buflen[, flags] ) | 从套接字接收数据，直到buflen字节，还返回数据来自的远程主机和端口 |
| sock.send( data[, flags] ) | 通过套接字发送数据 |
| sock.sendto( data[, flags], addr ) | 通过套接字发送数据，根据指定的地址 |
| sock.close() | 关闭套接字 |
| sock.getsockopt( lvl, optname ) | 获取指定套接字选项的值 |
| sock.setsockopt(lvl, optname, val ) | 设置指定套接字选项的值 |

类方法和实例方法的不同之处在于，实例方法需要执行一个套接字实例（从套接字返回），而类方法则不需要。

## 3.3 SocketServer

SocketServer模块是一个有趣的模块，可以简化套接字服务器的开发。 关于它的使用的讨论远远超出了本教程的范围，但是我将演示它的基本用法，点击[这个链接地址](https://www.ibm.com/developerworks/linux/tutorials/l-pysocks/#resources)，以获得更详细讨论的链接。

例2，简单示例。在这里，我实现了一个简单的“Hello World”服务器，它在客户端连接时发出消息。 我首先创建一个继承了SocketServer.StreamRequestHandler类的请求处理程序。 我定义了一个称为handle的方法来处理服务器的请求。 服务器所做的一切必须在这个函数的上下文中处理（最后，套接字被关闭）。 这个过程很简单，但是你可以用这个类来实现简单的HTTP服务器。 在处理方法中，我发出我的招呼，然后退出。

现在连接处理程序已准备就绪，剩下的就是创建套接字服务器。我使用  
SocketServer.TCPServer类，提供地址和端口号（服务器将绑定到）和我的请求处理程序方法。得到的是一个TCPServer对象。 调用serve\_forevermethod启动服务器并使其可用于连接。

例2. 用SocketServer模块实现一个简单的服务器

import SocketServer

class hwRequestHandler(SocketServer.StreamRequestHandler):

def handle(self):

self.wfile.write("Hello World!\n")

server = SocketServer.TCPServer ( ("", 2525), hwRequestHandler )

server. serve\_forever()

就是这样！ Python允许在这个主题上有许多变体，包括UDPServers，进程和线程服务器。

# 4 Python套接字编程

在使用套接字的语言中，套接字通常是相同的 - 它是两个可以相互通信的应用程序之间的通道。

## 4.1 依赖条件

无论你是用Python，Perl，Ruby，Scheme还是其他有用的语言编写套接字应用程序（并且我的意思是有套接字接口的语言），套接字都是相同的。 这是两个应用程序之间的通道，可以相互通信（本地在单个机器上，或者在两个不同位置的机器之间）。

与像Python这样的语言编程的套接字的区别在于可以简化套接字编程的辅助类和方法。 在本节中，我将演示Python套接字API。 您可以使用脚本执行Python解释器，或者如果您自己执行Python，则可以一次一行地与它交互。 这样，你可以看到每个方法调用的结果。

以下示例说明了与Python解释器的交互。 在这里，我使用套接字类方法gethostbyname将完全限定的域名（www.baidu.com）解析为字符串四段的IP地址（'61.135.169.125'）：

[camus]$ python  
Python 2.4 (#1, Feb 20 2005, 11:25:45)  
[GCC 3.2.2 20030222 (Red Hat Linux 3.2.2-5)] on linux2  
Type "help", "copyright", "credits" or "license" for more  
information.  
>>> import socket  
>>> socket.gethostbyname('www.baidu.com')  
'61.135.169.125'  
>>>

在导入套接字模块后，我调用gethostbyname类方法将域名解析为IP地址。

现在，我将讨论基本套接字方法和通过套接字进行通信。带着你的python自由跟随吧。

## 4.2创建和销毁套接字

要创建一个新的套接字，可以使用套接字类的套接字方法。 这是一个类方法，因为您还没有从中应用方法的套接字对象。 套接字方法与BSD API相似，如创建一个流（TCP）和数据报（UDP）套接字所示：

例4.创建流和数据报套接字

streamSock = socket.socket(socket.AF\_INET, socket.SOCK\_STREAM )  
dgramSock = socket.socket ( socket.AF\_INET, socket.SOCK\_DGRAM )

在每种情况下，返回一个套接字对象。 AF\_INET符号 - 参数1表示您正在请求Internet协议（IP）套接字，特别是IPv4。 第二个参数是传输协议类型（SOCK\_STREAM用于TCP套接字，SOCK\_DGRAM用于UDP套接字）。 如果您的底层操作系统支持IPv6，则还可以指定socket.AF\_INET6来创建IPv6套接字。

要关闭连接的套接字，请使用close方法：

streamSock.close()

最后，你可以用del语句删除一个套接字：

del streamSock

该语句永久删除套接字对象。 此后尝试引用套接字会产生错误。

## 4.3套接字地址

套接字的端点地址是由接口地址和端口号组成的元组。 因为Python可以很容易地表示元组，所以地址和端口也就很容易表示。这说明接口地址192.168.1.1和端口80的端点：

( '192.168.1.1', 80 )

您也可以在此处使用完全限定的域名，例如：

( 'www.ibm.com', 25 )

这个例子很简单，肯定会击败C中需要的sockaddr\_in操作。下面的讨论提供了Python中地址的例子。

## 4.4服务器套接字

服务器套接字通常是在网络上公开服务的套接字。由于服务器和客户端套接字是以不同的方式创建的，我将独立讨论它们。

创建套接字后，您使用bind方法将地址绑定到它，使用listen方法将其置于侦听状态，最后使用accept方法接受新的客户端连接。如下所示：

例5.使用服务器套接字

sock = socket.socket ( socket.AF\_INET, socket.SOCK\_STREAM )  
sock.bind(('', 2525) )  
sock.listen( 5 )  
newsock, (remhost, remport) = sock.accept()

对于这个服务器，使用地址（''，2525），这意味着通配符用于接口地址（''），允许来自主机上任何接口的传入连接。 您还绑定到端口号2525。

请注意，accept方法不仅返回表示客户端连接（newsock）的新套接字对象，还返回地址元组（套接字对端的远程地址和端口号）。 Python的SocketServer模块可以进一步简化这个过程，如上所示。

您也可以创建数据报服务器，但它们是无连接的，因此没有关联的接受方法。 以下示例创建一个数据报服务器套接字：

例6.创建一个数据报服务器套接字

sock = socket.socket( socket.AF\_INET, socket.SOCK\_DGRAM )  
sock.bind( ('', 2525) )

即将到来的关于套接字I/O的讨论显示了I/O如何适用于流和数据报套接字。

现在，我们来探讨客户端如何创建套接字并将其连接到服务器。

## 4.5 客户端套接字

创建和连接客户端套接字的机制与服务器套接字的设置相似。 在创建套接字时，需要一个地址 - 不要在本地绑定套接字（就像服务器一样），而是要确定套接字应该附加到哪里。 假设主机上有一台服务器，其接口IP地址为“192.168.1.1”，端口为2525.以下代码创建一个新的套接字并将其连接到定义的服务器：

例7.创建一个流套接字并连接到服务器

sock = socket.socket( socket.AF\_INET, socket.SOCK\_STREAM)  
sock.connect(('192.168.1.1', 2525))

对于数据报套接字，过程有点不同。 回想一下，数据报套接字本质上是断开的。 一种思考的方法如下：尽管流套接字是两个端点之间的管道，但是数据报套接字是基于消息的，能够同时与多个对等体进行通信。 这是一个数据报客户端的例子。

例8.创建一个数据报套接字并连接到服务器

sock = socket.socket( socket.AF\_INET, sock.sock\_DGRAM )  
sock.connect( ('192.168.1.1', 2525) )

这里的不同之处在于，即使我使用了连接方法，客户端和服务器之间也没有真正的连接。 这里的连接是对后面的I/O的简化。通常在数据报套接字中，您必须提供您要发送的数据的目标信息。通过使用连接，我已经与客户端缓存这个信息和发送方法可以发生很像流套接字版本（没有目标地址是必要的）。 您可以再次调用connect来重新指定数据报客户端消息的目标。

## 4.6 流套接字I/O

在Python中通过流套接字发送或接收数据很简单。存在几种通过流套接字移动数据的方法（如发送，接收，读取和写入）。

这第一个例子演示了一个服务器和客户端的流套接字。在这个演示中，服务器回应从客户端收到的任何内容。

echo服务器如例9所示。创建一个新的流套接字时，绑定一个地址（接受来自任何接口和端口45000的连接），然后调用listen方法来启用传入连接。echo服务器然后进入客户端连接的循环。 accept方法被调用，并阻塞（即不返回），直到新客户端连接，此时新客户端套接字与远程客户端的地址信息一起返回。使用这个新的客户端套接字，我调用recv从对端获取一个字符串，然后将这个字符串写回到套接字。然后，立即关闭了这次连接。

例9.简单的Python echo服务器

import socket

srvsock = socket.socket ( socket.AF\_INET, socket.SOCK\_STREAM )  
srvsock.bind( ('', 23000) )  
srvsock.listen( 5 )  
while 1:  
 clisock, (remhost, remport) = srvsock.accept()  
 str = clisock.recv(100)  
 clisock.send( str )  
 clisock.close()

例10显示了与例9中的服务器相对应的echo客户机。创建一个新的流套接字时，使用connect方法将此套接字连接到服务器。 当连接时（当连接方法返回时），客户端用send方法发出一个简单的文本消息，然后用recv方法等待回声。 print语句用于发出所读的内容。 完成后，执行close方法关闭套接字。

例10.简单的Python echo服务器

import socket

clisock = socket.socket( socket.AF\_INET, socket.SOCK\_STREAM )

clisock.connect( ('', 23000) )

clisock.send("Hello World\n")

print clisock.recv(100)

clisock.close()

## 4.7 数据报套接字I/O

数据报套接字本质上是断开的，这意味着通信要求提供目的地址。同样，当通过套接字接收到消息时，必须返回数据的来源。recvfrom和sendto方法支持额外的地址信息，正如您在数据报回显服务器和客户端实现中所看到的那样。

例11显示了数据报echo服务器。首先创建一个套接字，然后使用bind方法绑定到一个地址。 然后服务器进入一个无限循环，等待客户端的请求。recvfrom方法接收来自数据报套接字的消息，不仅返回消息，而且还返回消息源的地址。然后用sendto方法将这些信息转换回来，再发送给客户端。

例11. 简单的Python数据报echo服务器

import socket

dgramSock = socket.socket(socket.AF\_INET, socket.SOCK\_DGRAM )

dgramSock.bind(('', 23000) )

while 1:

msg, (addr, port) = dgramSock.recvfrom(100)

dgramSock.sendto(msg,(addr, port) )

数据报客户端更简单。创建数据报套接字后，我使用sendto方法将消息发送到特定的地址。 （记住：Datagrams没有连接。）sendto完成后，我用recv等待echo回应，然后打印它。 请注意，我不会在这里使用recvfrom，因为我对对等地址信息不感兴趣。

例12.简单Python数据报echo客户端

import socket

dgramSock = socket.socket( socket.AF\_INET, socket.SOCK\_DGRAM )  
dgramSock.sendto( "Hello World\n", ('', 23000) )

print dgramSock.recv( 100 )

dgramSock.close()

## 4.8 套接字选项

套接字默认为一组标准行为，但可以使用选项来改变套接字的行为。 您可以使用setsockopt方法操作套接字选项，并使用getsockopt方法捕获它们。

在Python中使用套接字选项很简单，如例13所示。在第一个示例中，我读取套接字发送缓冲区的大小。 在第二个例子中，我得到SO\_REUSEADDR选项的值（在TIME\_WAIT周期内重新使用该地址），然后启用它。

例13. 使用Socket套接字

sock = <strong>socket.socket(socket.AF\_INET, socket.SOCK\_STREAM )

# 获取socket发送缓存的大小

bufsize = sock.<strong>getsockopt( socket.SOL\_SOCKET, socket.SO\_SNDBUF )

# 获取SO\_REUSEADDR选项

state = sock. getsockopt ( socket.SOL\_SOCKET, socket.SO\_REUSEADDR )

# 使能SO\_REUSEADDR选项

sock.setsockopt( socket.SOL\_SOCKET, socket.SO\_REUSEADDR, 1 )

SO\_REUSEADDR选项最常用于套接字服务器开发。您可以增加套接字发送和接收缓冲区以获得更好的性能，但是鉴于您使用的是解释型脚本语言，因此可能无法为您带来太多好处。

## 4.9 异步I/O

Python提供异步I / O作为select模块的一部分。这个功能类似于C select机制，但有一些简化。我将首先介绍select，然后向您展示如何在Python中使用它。

select方法允许您监听多个套接字的复用事件和几个不同的事件。例如，您可以指示select在套接字有数据可用时通知您，可以通过套接字写入数据时以及在套接字上发生错误时通知您; 您可以同时为多个套接字执行这些操作。

当C用位工作时，Python使用列表来表示要监视的描述符以及其约束满足的返回描述符。 考虑下面的例子，你等待来自标准输入的一些输入：

例14. 等待来自stdin的输入

rlist, wlist, elist = select.select( [sys.stdin], [], [] )

print sys.stdin.read()

传递给select的参数是表示读取事件，写入事件和错误事件的列表。 select方法返回三个包含事件满足的对象的列表（读，写，异常）。 在这个例子中，返回rlist时应该是[sys.stdin]，表明数据可以在stdin上读取。 然后用读取方法读取这些数据。

select方法也适用于套接字描述符。在下面的例子中（见清单15），创建了两个客户端套接字并连接到远程对等端。 然后使用select方法来确定哪个套接字有数据可供读取。 数据然后被读取并发送到stdout。

例15. 用多个套接字演示select方法

import socket

import select

sock1 = socket.socket( socket.AF\_INET, socket.SOCK\_STREAM )

sock2 = socket.socket( socket.AF\_INET, socket.SOCK\_STREAM )

sock1.connect(('192.168.1.1', 25) )

sock2.connect(('192.168.1.1', 25) )

while 1:

# 等待一个read事件

rlist, wlist, elist = select.select([sock1,sock2], [], [], 5 )

# 测试timeout

if [rlist, wlist, elist] == [ [], [], [] ]:

print "Five seconds elapsed.\n"

else:

# 遍历rlist列表中的每一个Socket, 读取并打印可用的数据

for sock in rlist:

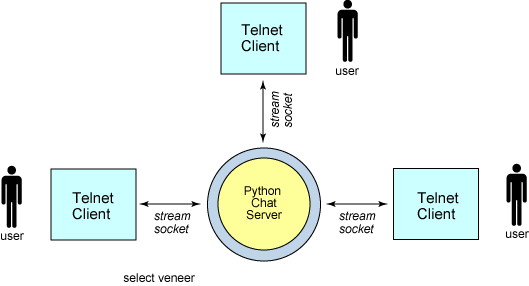
print sock.recv(100)

# 5构建一个Python聊天服务器

## 5.1 一个简单的聊天服务器

您已经探索了Python的基本网络API;现在你可以把这些知识用在一个简单的应用程序中。 在本节中，您将构建一个简单的聊天服务器。使用Telnet，客户端可以连接到您的Python聊天服务器，并在全服务器范围内相互通信。提交给聊天服务器的消息被其他人查看（除了管理信息，例如客户端加入或离开聊天服务器）。该模型如图1所示。

图1. 聊天服务器使用select方法来支持任意数量的客户端



对于这个聊天服务器，一个重要的要求是它必须是可扩展的。服务器必须能够支持任意数量的流（TCP）客户端。

要支持任意数量的客户端，请使用select方法异步管理客户端列表。但是你也使用select的特性来作为你的服务器套接字。select的select事件决定了客户端何时有数据可读，但是也可以用来确定服务器套接字何时有新客户端尝试连接。 您利用此行为来简化服务器的开发。

接下来，我将探索Python聊天服务器的源代码，并确定Python如何帮助简化实现。

## 5.2 ChatServer类

我们先看看Python聊天服务器类和\_\_init\_\_方法 - 创建新实例时调用的构造函数。

该类由四个方法组成。run方法被调用来启动服务器并允许客户端连接。 broadcast\_string和accept\_new\_connection方法在类中内部使用，稍后将会讨论。

\_\_init\_\_方法是在创建类的新实例时调用的特殊方法。请注意，所有方法都可以self参数，即对类实例本身的引用（非常类似于C ++中的this参数）。 您将看到self参数，这是所有实例方法的一部分，用于访问实例变量。

\_\_init\_\_方法创建三个实例变量。port是服务器的端口号（在构造函数中传递）。srvsock是此实例的套接字对象，描述符是包含该类的每个套接字对象的列表。您可以在select方法中使用此列表来标识读取事件列表。

最后，例16显示了\_\_init\_\_方法的代码。创建一个流套接字之后，启用SO\_REUSEADDR套接字选项，以便服务器可以在必要时快速重新启动。通配符地址与定义的端口号绑定。然后调用listen方法来允许即将接入的连接。服务器套接字被添加到描述符列表（目前唯一的元素），但所有客户端套接字将在到达时添加（请参阅accept\_new\_connection方法）。标准输出提供给标准输出，表示服务器已经启动。

例16. 带有init 方法的ChatServer类

import socket

import select

class ChatServer:

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

self.port = port;

self.srvsock = socket.socket(socket.AF\_INET, socket.SOCK\_STREAM )

self.srvsock.setsockopt(socket.SOL\_SOCKET, socket.SO\_REUSEADDR, 1 )

self.srvsock.bind( ("", port) )

self.srvsock.listen(5)

self.descriptors = [self.srvsock]

print 'ChatServer started on port %s' % port

def run( self ):

...

def broadcast\_string( self, str, omit\_sock ):

...

def accept\_new\_connection( self ):

## 5.3 run方法

run方法是聊天服务器的主循环（参见例17）。当被调用时，它进入一个无限循环，提供连接的客户端之间的通信。

服务器的核心是select方法。我将描述符列表（包含所有服务器的套接字）作为读取事件列表传给select函数（对于写入和异常，列表为空列表）。当检测到读取事件时，sread列表将会被返回（这里，我忽略swrite和sexc列表）。sread列表包含将被服务的套接字对象。我遍历sread列表，检查每个找到的套接字对象。

在主循环中，首先检查套接字对象是否是服务器。如果是，则新客户端正在尝试连接，调用accept\_new\_connection方法。否则，读取客户端套接字。如果从recv返回null，则关闭对等套接字。

在这种情况下，我构建一个消息并将其发送给所有连接的客户端，关闭对等套接字，并从描述符列表中删除相应的对象。如果recv返回值不为空，则消息可用并存储在str中。此消息使用broadcast\_string分发给所有其他客户端。

例17. 聊天服务器run方法是聊天服务器的核心

def run( self ):

while 1:

# 等待一个可读的套接字描述符的事件

(sread, swrite, sexc) = select.select( self.descriptors, [], [] )

# 遍历标记的可读描述符

for sock in sread:

# 收到一个连接到服务器（侦听）套接字

if sock == self.srvsock:

self.accept\_new\_connection()

else:

# Received something on a client socket

str = sock. recv (100)

# Check to see if the peer socket closed

if str == '':

host,port = sock.getpeername()

str = 'Client left %s:%s\r\n' % (host, port)

self.broadcast\_string( str, sock )

sock.close

self.descriptors.remove(sock)

else:

host,port = sock.getpeername()

newstr = '[%s:%s] %s' % (host, port, str)

self.broadcast\_string( newstr, sock )

## 5.4 Helper方法

聊天服务器类中的两个辅助方法提供了接受新的客户端连接和广播消息给连接的客户端的方法。

当在等待连接队列中检测到新的客户端，就会调用accept\_new\_connection方法（参见例18）。accept方法用于接受连接，它将返回新的套接字对象和远程地址信息。立即添加新的socket到descripters列表中。我创建了一个字符串，标识客户端已经连接并使用broadcast\_string方法将这些信息广播到组（参见例19）。

注意除了要被广播的字符串，socket对象也会被传递。原因是我想有选择地忽略一些套接字避免它们得到某些特定信息。例如，当一个客户端发送一条消息给该组时，消息被广播给组，但是不应该再发给它自己本身。还比如，当我产生标志新的客户端加入组的身份信息时，这条信息应该被广播到组，但是不应该包含新的客户端自身。这个功能的实现，就是在broadcast\_string方法中使用omit\_sock参数。此方法遍历描述符列表并将该字符串发送到所有不是服务器套接字和notomit\_sock的套接字。

例18. 在聊天服务器上接受新的客户端连接

def accept\_new\_connection( self ):

newsock, (remhost, remport) = self.srvsock.accept()

self.descriptors.append(newsock)

newsock.send("You're connected to the Python chatserver\r\n")

str = 'Client joined %s:%s\r\n' % (remhost, remport)

self.broadcast\_string( str, newsock )

例19. 将消息广播到聊天组

def broadcast\_string( self, str, omit\_sock ):

for sock in self.descriptors:

if sock != self.srvsock and sock != omit\_sock:

sock.send(str)

print str,

## 5.5实例化一个新的ChatServer

现在你已经看到了Python聊天服务器（在50行以下的代码），让我们看看如何在Python中实例化一个新的聊天服务器对象。

通过创建一个新的ChatServer对象启动服务器（别忘了传递port！），然后调用run方法来启动服务器并允许传入的连接：

例20. 实例化一个聊天服务器

myServer = ChatServer( 2626 )

myServer.run()

至此，服务器已经运行且你可以使用一个或者多个客户端连接它。您也可以将方法链接在一起来简化这个过程（就像它需要更简单一样）：

例21. 组合

myServer = ChatServer( 2626 ).run()

它也可以达到相同的结果。

## 5.6 验证ChatServer

这是ChatServer实际运行的效果。我将展示ChatServer服务器的输出（参见例22）和2个客户端的对话过程（参见例23和例24）。用户输入的文本以粗体显示。

例22.ChatServer服务器的输出

[plato]$ **python pchatsrvr.py**

ChatServer started on port 2626

Client joined 127.0.0.1:37993

Client joined 127.0.0.1:37994

[127.0.0.1:37994] Hello, is anyone there?

[127.0.0.1:37993] Yes, I'm here.

[127.0.0.1:37993]  Client left 127.0.0.1:37993

例23. 聊天客户端#1的输出

[plato]$ **telnet localhost 2626**

Trying 127.0.0.1...

Connected to localhost.localdomain (127.0.0.1).

Escape character is '^]'.

You're connected to the Python chatserver

Client joined 127.0.0.1:37994

[127.0.0.1:37994] Hello, is anyone there?

**Yes, I'm here.**

telnet> close

Connection closed.

[plato]$

例 24. 聊天客户端#2的输出

[plato]$ **telnet localhost 2626**

Trying 127.0.0.1...

Connected to localhost.localdomain (127.0.0.1).

Escape character is '^]'.

You're connected to the Python chatserver

**Hello, is anyone there?**

[127.0.0.1:37993] Yes, I'm here.

[127.0.0.1:37993] Client left 127.0.0.1:37993

如例22所示，所有客户端之间的所有对话都被发送到stdout，包括客户端连接和断开连接消息。

# 6 高级网络类

## 6.1 网络模块

Python包括了多个专用的应用层协议的模块（基于标准socket模块构建）。可用模块种类繁多，提供了超文本传输协议（HTTP），简单邮件传输协议（SMTP），因特网消息访问协议（IMAP）和邮局协议（POP 3），网络新闻传输协议（NNTP），XML-RPC（远程过程调用），FTP等等。

本节介绍表4中显示的模块。

表4. 有用的应用层协议模块

|  |  |
| --- | --- |
| 模块 | 实施的协议 |
| httplib | HTTP客户端 |
| smtplib | SMTP客户端 |
| poplib | POP3客户端 |

## 6.2 httplib(HTTP 客户端)

开发Web机器人或其他Internet抓取代理时，HTTP客户端接口可能很有用。Web协议本质上是流套接字上的请求/响应。Python使得通过简单的Web界面构建Web机器人变得容易。

例25演示了httplib模块。使用HTTPConnection创建一个新的HTTP客户端实例，需要提供您想要连接的网站。 有了这个新的对象（httpconn），你可以用request方法请求文件。 在request请求里，您可以指定HTTP GET方法（从服务器请求文件，与HEAD相比，它只是简单地检索有关文件的信息）。 getresponse方法解析HTTP响应头，以了解是否返回错误。如果文件被成功检索，新的响应对象的读取方法返回并打印文本。

例25. 使用httplib构建一个简单的HTTP客户端(非渲染)

import httplib

httpconn = **httplib.HTTPConnection**("www.ibm.com")

httpconn.**request**("GET", "/developerworks/index.html")

resp = httpconn.**getresponse**()

if resp.reason == "OK":

resp\_data = resp.**read**()

print resp\_data

httpconn.**close**()

### 6.2.1 HTTPConnection类

1. HTTPConnection类的构造函数

语法：

httplib.HTTPConnection(host[,port[,strict[,timeout]]])

表示一次与服务器之间的交互，即请求/响应。

参数说明：

* host 表示服务器主机，如：[www.csdn.net](http://www.csdn.net)；
* port 为端口号，默认值为80；
* strict 默认值为false，表示在无法解析服务器返回的状态行时( status line)（比较典型的状态行如：HTTP/1.0 200 OK），是否抛BadStatusLine 异常；
* timeout 可选参数，表示超时时间。

2 HTTPConnection的方法：

（1）HTTPConnection.request ( method , url [ , body [ , headers ]] )

调用request 方法会向服务器发送一次请求。

参数说明：

* method 表示请求的方法，常用有方法有get 和post ；
* url 表示请求的资源的url ；
* body 表示提交到服务器的数据，必须是字符串（如果method 是”post” ，则可以把body 理解为html 表单中的数据）；
* headers 表示请求的http头

（2）HTTPConnection.getresponse()

获取Http响应。返回的对象是HTTPResponse的实例，关于HTTPResponse在下面会讲解。

（3）HTTPConnection.connect()

连接到Http 服务器。

（4）HTTPConnection.close ()

关闭与服务器的连接。

（5）HTTPConnection.set\_debuglevel(level)

设置高度的级别。参数level 的默认值为0 ，表示不输出任何调试信息。

### 6.2.2 httplib.HTTPResponse

表示服务器对客户端请求的响应。往往通过调用HTTPConnection.getresponse()来创建，它有如下方法和属性：

1. HTTPResponse.read([amt])

获取响应的消息体。如果请求的是一个普通的网页，那么该方法返回的是页面的html。可选参数amt表示从响应流中读取指定字节的数据。

2. HTTPResponse.getheader(name[, default])

获取响应头。Name表示头域(header field)名，可选参数default在头域名不存在的情况下作为默认值返回。

3. HTTPResponse.getheaders()

以列表的形式返回所有的头信息。

4. HTTPResponse.msg

获取所有的响应头信息。

5. HTTPResponse.version

获取服务器所使用的http协议版本。11表示http/1.1；10表示http/1.0。

6. HTTPResponse.status

获取响应的状态码。如：200表示请求成功。

7. HTTPResponse.reason

返回服务器处理请求的结果说明。一般为”OK”。

### 6.2.3 补充说明

1. 以上的说明是基于Python2.7版本来说的。在Python3中已经httplib模块改为http\_client模块了。

2. 现在如果想访问网站，比如说[www.ibm.com](http://www.ibm.com)，就要使用SSL安全证书，所以在httplib模块中对应的方式是httplib.HTTP**S**Connection(host)，那么对应的上面的示例代码应该修改如下：

import httplib

httpconn = **httplib.HTTPSConnection**("www.ibm.com")

httpconn.**request**("GET", "/developerworks/index.html")

resp = httpconn.**getresponse**()

if resp.reason == "OK":

resp\_data = resp.**read**()

print resp\_data

httpconn.**close**()

## 6.3 smptlib (SMTP客户端)

SMTP允许您将电子邮件发送到邮件服务器，该邮件服务器可以在网络系统中中转你的邮件。用于发送电子邮件的Python模块很简单，包括创建SMTP对象，使用sendmail方法发送电子邮件，然后使用quit方法关闭连接。

例26中的示例演示了发送一个简单的电子邮件消息。 msg字符串包含消息的正文（应该包含主题行）。

例26.用smtplib发送一封简短的电子邮件

#!/usr/bin/env python # start line

# coding:utf-8

import smtplib

from email.mime.text import MIMEText

mail\_host = "smtp.163.com" # 163邮箱发送服务器

mail\_user = "user" # 用户名

mail\_pass = "password" # 邮箱的授权密码，不是邮箱的登录密码

sender = 'sender\_user\_addr ' # 发送方邮箱地址

receivers = ['rcv\_user\_addr'] # 接收方邮箱地址

content = '123456!'

title = 'Python SMTP Mail Test'

message = MIMEText(content, 'plain', 'utf-8')

message['From'] = "{}".format(sender)

message['To'] = ",".join(receivers)

message['Subject'] = title

try:

smtpObj = smtplib.SMTP\_SSL(mail\_host, 465)

smtpObj.login(mail\_user, mail\_pass)

smtpObj.sendmail(sender, receivers, message.as\_string())

print("mail has been send successfully.")

except smtplib.SMTPException as e:

print(e)

## 6.4 poplib(POP3 客户端)

POP3是Python中另一个有用的应用程序层协议。POP3协议允许您连接邮件服务器并下载新邮件，这对于远程命令-在电子邮件正文中嵌入命令非常有用。执行嵌入式命令后，可以使用smptlib将响应电子邮件返回给源。

例27中，演示了一个简单的应用程序，该应用程序连接到一个邮件服务器，并为用户发出所有挂起电子邮件的主题行。

poplib比较简单，但是提供了几种在服务器上收集和管理电子邮件的方法。在这个例子中，我使用POP3方法创建了一个新的POP3对象，指定为邮件接收服务器。用户和pass\_方法验证应用程序到服务器; stat方法返回等待用户的消息数和所有消息占用的总字节数。

接下来，我遍历每个可用的消息并使用retr方法获取下一封电子邮件。这个方法返回一个表单的列表：

(response, ['line, ...], octets)

在这里，response是特定消息的POP3响应；line列表，表示电子邮件消息的各行；最后一个元素，octets表示电子邮件消息的字节数。里边的循环只是遍历电子邮件正文列表的第二个元素（下标为[1]）。对于每一行，我测试是否存在“Subject：”;如果是的话，打印这一行。

在检查完所有电子邮件的消息之后，调用quit方法，结束POP3会话。

您也可以使用top方法来提取电子邮件的标题，而不是使用retr方法。这一步将更快，并最大限度地减少传输到这个客户端的数据量。

例27. 从POP3邮件服务器检索电子邮件并发出主题行

#!/usr/bin/env python # start line

# coding:utf-8

import poplib

import re

mail\_host = "pop.163.com" # 163 邮箱发送服务器

mail\_user = "user" # 163 邮箱用户

mail\_pass = "password" # 163 邮箱授权密码，不是邮箱登录密码

popClient = poplib.POP3\_SSL(mail\_host) # SSL安全证书访问，与poplib.POP3一样

popClient.user(mail\_user)

popClient.pass\_( mail\_pass)

numMsgs, mboxSize = popClient.stat()

print "Number of messages ", numMsgs

print "Mailbox size", mboxSize

print

for id in range (numMsgs):

for mail in popClient.retr(id+1)[1]:

if re.search( 'Subject:', mail ):

print mail

print

popClient.quit()

### 6.4.1 poplib类

python的poplib模块是用来从pop3收取邮件的，也可以说它是处理邮件的第一步。POP3协议并不复杂，它也是采用的一问一答式的方式，你向服务器发送一个命令，服务器必然会回复一个信息。pop3命令码如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 命令 | poplib方法 | 参数 | 状态 | 描述 |
| USER | user | username | 认可 | 用户名，此命令与下面的pass命令若成功，则导致状态转换 |
| PASS | pass\_ | password | 认可 | 用户密码 |
| APOP | apop | Name,Digest | 认可 | Digest是MD5消息摘要 |
| STAT | stat | None | 处理 | 请求服务器发回关于邮箱的统计资料，如邮件总数和总字节数 |
| UIDL | uidl | [Msg#] | 处理 | 返回邮件的唯一标识符，POP3会话的每个标识符都将是唯一的 |
| LIST | list | [Msg#] | 处理 | 返回邮件数量和每个邮件的大小 |
| RETR | retr | [Msg#] | 处理 | 返回由参数标识的邮件的全部文本 |
| DELE | dele | [Msg#] | 处理 | 服务器将由参数标识的邮件标记为删除，由quit命令执行 |
| RSET | rset | None | 处理 | 服务器将重置所有标记为删除的邮件，用于撤消DELE命令 |
| TOP | top | [Msg#] | 处理 | 服务器将返回由参数标识的邮件前n行内容，n必须是正整数 |
| NOOP | noop | None | 处理 | 服务器返回一个肯定的响应 |
| QUIT | quit | None | 更新 |  |

Python的poplib也针对这些命令分别提供了对应的方法,上面在第二列里已经标出来。收取邮件的过程一般是：  
1. 连接pop3服务器 (poplib.POP3.\_\_init\_\_)  
2. 发送用户名和密码进行验证 (poplib.POP3.user poplib.POP3.pass\_)  
3. 获取邮箱中信件信息 (poplib.POP3.stat)  
4. 收取邮件 (poplib.POP3.retr)  
5. 删除邮件 (poplib.POP3.dele)  
6. 退出 (poplib.POP3.quit)

注意的是，上面我在括号里写的是使用什么方法来完成这个操作，在实际的代码中不能那样写，应该是创建poplib.POP3的对象，然后，调用这个对象的方法。比如：  
poplib.POP3.quit  
应该理解为  
a = poplib.POP3(host)  
a.quit()

# 7总结

本教程回顾了套接字API的基础知识，并展示了如何使用Python构建网络应用程序。引入了标准套接字模块作为构建客户端和服务器应用程序的一种方式，以及简化了简单套接字服务器构建的SocketServer模块。我介绍了一个用Python实现的简单聊天服务器，它使用select方法为可扩展的客户端提供支持。最后，我预览了一些Python的高级网络类，它们简化了需要应用层网络协议的应用程序的开发。

Python是一个有趣而有用的语言，非常值得您学习。