

Python 新员工教材

楚广明 2012

新浪微博：@疯狂制造

请关注我，以便在第一时间获取更新消息

借用了一些闫小勇\郑纳智\学无止境等等同志的文档，向同志们致敬！

因为时间久远，很多引用文档以找不到作者，如果您发现我引用了您的文档，请第一时间联系我
引用作者目录我放在以下网址：<http://2.myrestful.sinaapp.com/>

目录

一、	Python 概述	8
1)	第一个 Python 程序	8
1.	为什么要学习 Python	10
2.	代码块与缩进的差异	10
3.	语句结尾的差异	10
4.	注释方法	10
5.	入口方法	11
6.	import 与 using 语句	11
2)	小结	11
二、	基本数据类型	11
1)	第一个程序的解析	11
2)	变量的命名规范	13
3)	空类型	14
4)	布尔类型	14
5)	数值类型	14
6)	字符串类型	15
7)	全局变量	16
三、	基本数据类型-列表(list【】)	16
1)	概述	16
2)	常用方法	18
四、	基本数据类型-字典(dict {})	19
1)	概述	19
2)	常用方法	21
五、	基本数据类型-元组(tuple ())	21
六、	基本数据类型-集合(set)	23
七、	数据类型转换	23
1)	概述	23
2)	常用方法	24
八、	运算符、表达式和流程控制	24
1)	算术运算符与算术表达式	24
2)	赋值运算符与赋值表达式	25
3)	关系运算符与关系表达式	25
九、	流程控制	26
1)	条件语句	26
2)	循环	26
十、	异常	27
十一、	动态表达式	28
十二、	函数及函数编程	29
1)	函数的定义	29
2)	定义一个交换函数	30
3)	函数的默认参数与返回值	30
4)	返回多个值	30

5)	locals 函数和 globals 函数介绍	31
十三、	类及面向对象	31
1)	类的定义	32
2)	为类添加数据	32
3)	构造函数	32
4)	静态成员与私有成员	33
5)	为类添加方法	33
6)	静态方法	34
7)	单继承	35
8)	传值与传引用	36
十四、	模块与包	37
1)	对于模块的理解	37
2)	模块的显要特性：属性	38
3)	模块创建过程的例子	38
4)	模块的导入	39
5)	模块的属性	39
十五、	字符串与简单的正则表达式	39
1)	解决中文字符的问题	40
2)	字符串的格式化	41
3)	字符串的合并与分割	42
4)	常用字符操作演示	44
十六、	文件 I/O	46
1)	基本文件功能演示	46
2)	文件的写入	48
3)	文件的删除与复制	48
4)	文件与目录的重命名	49
5)	文件内容的查找和替换	50
6)	文件的比较	51
7)	控制台输入	51
十七、	中文乱码	51
1)	编码普及	52
1.	ASCII 码	52
2.	非 ASCII 编码	52
3.	Unicode	52
4.	Unicode 的问题	52
5.	UTF-8	53
6.	Unicode 与 UTF-8 之间的转换	53
7.	Little endian 和 Big endian	54
2)	中文乱码细究	55
1.	python 中的 str 和 unicode	55
2.	str 和 unicode 对象的转换	55
3)	Setdefaultencoding	56
3.	操作不同文件的编码格式的文件	56
4.	文件的编码格式和编码声明的作用	57
4)	新手编码必读（gonewiththewind 大虾作品）	58
十八、	Web.Py 框架概述	59
1)	Python 下 web 开发框架的选择	59
1.	Django	59

2.	Pylons & TurboGears & repoze.bfg.....	60
3.	Tornado & web.py	61
4.	Bottle & Flask.....	61
5.	Quixote.....	62
6.	最后关于框架选择的误区.....	62
2)	解决 SAE 第三方包的安装方案	62
3)	web.py 安装	64
4)	web.py 下常用框架简介	65
5)	最简单的 HelloWorld!.....	65
6)	Web.py 基本使用模板使用	65
7)	进阶模板使用	67
1.	server2.py	67
2.	post.html	67
8)	URL 控制	68
1.	问题:	68
2.	解决:	68
3.	路径匹配	68
9)	一个完整的小例子	68
10)	web.seeother 和 web.redirect 转向.....	69
1.	问题	69
2.	解法	69
3.	区别	69
11)	包含应用	70
1.	实现	70
12)	使用 XML	70
1.	问题	70
2.	解法	71
13)	获取 POST 数据	71
1.	login.html 模板	72
2.	主程序	72
14)	获取客户端信息	73
1.	问题	73
2.	解法	73
3.	例子	73
4.	'ctx' 中的数据成员	73
15)	文件上传	74
1.	简单的文件上传	74
2.	基于新浪云的上传	74
十九、	Web.py 使用 Session	76
1)	问题	76
2)	解法	76
3)	在 template 下使用 Session	78
1.	问题:	78
2.	解决:	78
4)	如何操作 Cookies	79
1.	问题	79
2.	解法	79
3.	设置 Cookies	79

4.	获得 Cookies.....	80
5)	布局模板	81
1.	问题	81
2.	方法	81
3.	Tip: 在布局文件（layout.html）中定义的页面标题变量	82
4.	Tip: 在其他模板中引用 css 文件，如下：	82
二十、	Web.py 支持的语法模板	83
1)	使用 Templeteor 模板语法	83
1.	Introduction.....	83
2.	使用模板系统	83
3.	表达式用法.....	84
4.	赋值	84
5.	过滤	85
6.	新起一行用法	85
7.	转义 \$	86
8.	注释	86
9.	控制结构	86
10.	使用 def	87
11.	使用 var	87
12.	内置 和 全局	88
2)	使用 Jinja2 模板语法	89
1.	上手的一个小例子	89
二十一、	Web.py 中使用数据库	90
1)	使用 web.py 自带的 DB 封装库	90
1.	一个简单的数据库访问测试	90
2.	一个复杂一点的例子	91
3.	更加复杂的例子	91
二十二、	使用 Sqlalchemy	93
1)	简介	93
2)	初始化 Sqlite 数据库	95
3)	插入数据库	96
4)	在视图中使用 sqlalchemy	99
1.	首先创建数据库 models.py	99
2.	使用 app.py.....	100
二十三、	Json 的使用	101
二十四、	Urllib 模块	102
1)	urlopen 的基本使用	102
2)	urlretrieve 方法的使用	103
3)	url 编码	103
4)	使用 httpplib 抓取	104
二十五、	Urllib2 的使用	106
1)	最简单的爬虫	106
2)	提交表单数据	106
1.	用 GET 方法提交数据	106
二十六、	操作 Memcached	107
1)	安装	107
1.	Linux 环境	107
2.	Win 环境.....	107

3.	linux 启动 memcached	107
4.	windows 下启动.....	108
2)	Python 操作 Memcached	108
1.	Python 操作 memcached	109
2.	Memcached 常用方法	109
二十七、	SQLite 模块	109
1)	简单的介绍	109
2)	安装与使用	110
3)	一个简单的例子.....	112
4)	中文处理	114
二十八、	正则表达式	114
1)	简介.....	114
2)	3 个重要的正则式命令	115
1.	验证字符串.....	117
2.	验证 url.....	117
3.	验证 Email.....	118
4.	验证值在 0-25 的数字	118
5.	验证格式为 MM/DD/YYYY, YYYY/MM/DD and DD/MM/YYYY 的日期	118
3)	正则表达符	119
1.	你常用的元字符	119
2.	常用的限定符	119
3.	字符串类	120
4.	分枝条件	121
5.	分组	121
6.	反义	122
7.	后向引用	122
8.	贪婪与懒惰	123
4)	在 Python 中使用	123
5)	Re 模块	125
6)	Match 方法	126
7)	Pattern 方法	128
二十九、	Python 中的进程与线程	133
1)	进程和线程	133
2)	进程的运行环境.....	133
3)	创建进程	133
1.	system 函数.....	133
三十、	Pyexe	134
三十一、	Sina SDK 的使用	135
1)	首先申请 SDK 开发者.....	135
2)	下载 Python SDK	137
3)	一个简单的演示程序	137
4)	一个进阶的演示.....	138
三十二、	附录-Ubuntu 开发环境之安装	140
1)	进入 live cd 桌面	140
2)	安装系统	142
3)	连网换源	149
4)	更新系统	154
三十三、	附录 vmware 中的 bridge、nat 区别	159

1)	VMware 网络连接的几种方式	159
2)	replicate physical network connection state 选项解释:	163
三十四、	附录 Python 学习资源大全	164
1)	一、学会使用搜索引擎，尤其是 Google。推荐阅读文章:	164
2)	二、学会提问	164
3)	三、书籍推荐:	164
4)	四、视频推荐:	164
5)	五、学习方法	164
1.	知乎问答：零基础，应当如何开始学习 Python ？点击打开链接--by@黎敏	164
2.	芝麻问答：如何学习 Python 点击打开链接作者：@halida	166
3.	知乎问答：你是如何自学 Python 的？点击打开链接 作者:@成增存	167
6)	六、Just Do It.....	167
7)	七、更新.....	167

一、 Python 概述

Python 是一门优雅而健壮的编程语言，它继承了传统编译语言的强大性与通用性，同时也借鉴了简单的脚本和解释语言的易用性。它可以帮你完成工作，而且一段时间之后，你还能看明白自己写的这段代码。你会对自己如此快地学会它和它强大的功能感到十分的惊讶，更不用提你已经完成了工作了，只有你想不到，没有 **Python** 做不到的。

就算你的项目中有大量的 **Python** 代码，你也依旧可以有条不紊地通过将其分离为多个文件或模块加以组织管理。而且你可以从一个模块中选取代码，而从另一个模块中读取属性。更棒的是，对于所有模块，**Python** 的访问语法都是相同的。不管这个模块是 **Python** 标准库中的还是你一分钟前创建的，哪怕是你用其他语言写的扩展都没有问题！借助这些特点，你会自己根据需要扩展了这门语言，而且你也这么做了。

代码中的瓶颈可能是在性能分析中总排在前面的那些热门或者一些特别强调性能的地方，可以作为 **Python** 扩展用 C 重写。需要重申的是，这些接口和纯 **Python** 模块的接口是一模一样的，乃至代码和对象的访问方法也如出一辙的，唯一不同的是，这些代码为性能带来了显著的提升，我们可以利用 Pyrex 这样的工具允许 C 和 **Python** 混合编程，使编写扩展轻而易举，因为它会把所有的代码都转换成 C 语言代码。

因为 **Python** 的标准实现是使用 C 语言完成的（也就是 CPython），所以要使用 C 和 C++ 编写 **Python** 扩展。**Python** 的 java 实现被称为 Jython，要使用 java 编写其扩展。最后还有 IronPython 这是针对 .net 平台的实现。

在各种不同的系统上你都可以看到 **Python** 的身影，因为 **Python** 是用 C 写的，又由于 C 的可移植性，使得 **Python** 可以运行在任何带有 ANSI C 编译平台上。

- 内置的数据类型

Python 提供了一些内置的数据结构，这些数据结构实现了类似 Java 中集合类的功能，**Python** 的数据结构包括元组、列表、字典等。内置的数据结构简化了程序的设计。元组相当于“只读”的数组，列表可以作为可变长度的数组使用，字典相当于 java 中的 HashTable 类型。

- 健壮性

Python 提供了异常处理机制，能捕获程序的异常情况。此外，**Python** 的堆栈跟踪对象能够指出程序出错的位置和出错的原因。异常机制能够避免不安全退出的情况，同时能帮助程序员调试程序。

- 跨平台性

Python 会先编译与平台相关的二进制代码，然后再解释执行，这种方式和 Java 相似。**Python** 可以运行在 Windows/Linux/MAC/Unix 上

- 可扩展性

Python 是采用 C 开发的语言，因此可以使用 C 扩展 **Python**，可以给 **Python** 添加新的模块、新的类。

- 动态性

Python 与 Javascript、PHP、Perl 等语言类似。**Python** 不需要声明变量，直接赋予值可创建一个新的变量。

- 强类型语言

Python 的变量创建后会对应一种数据类型，**Python** 会根据赋值表达式的内容决定变量的数据类型。**Python** 在内部建立了管理这些变量的机制，出现在同一个表达式中的不同类型的变量需要做类型转换。

- 应用广泛

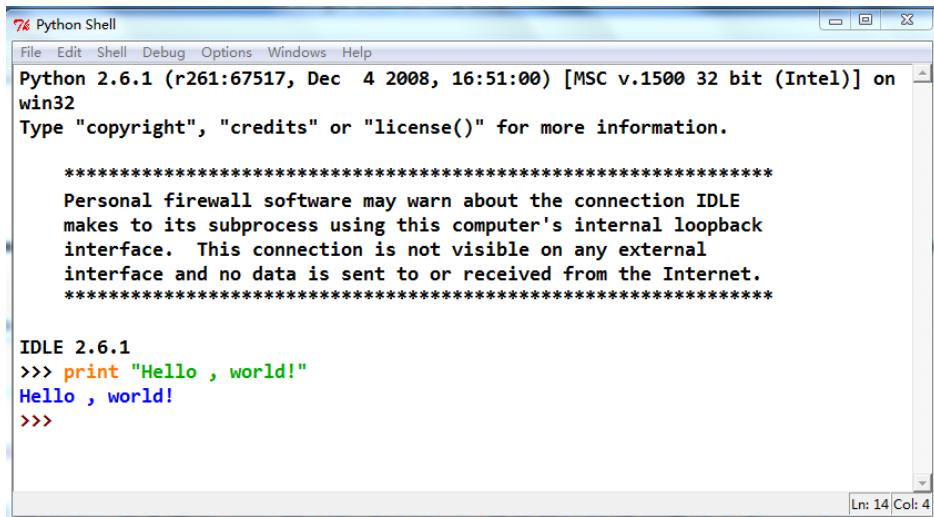
Python 语言应用于数据库、网络、图形图像、数学计算、WEB 开发、操作系统扩展等领域。**Python** 有许多第三方库的支持。例如，PIL 库用于图像处理、NumPy 库用于数据计算、WxPython 用于 GUI 库的设计、Django 库用于 WEB 应用程序的开发等

1) 第一个 Python 程序

在 **Python** 的官方网站可以下载到 Windows 下的安装包，按照提示一路下去就可以了，记得要将 **Python** 所在的目录加入到系统 Path 变量中。

Python 的安装包自带了一个简单的集成开发环境 **IDLE**，你也可以选一个自己喜欢的 **IDE**，我个人推荐 **PythonWin**，它的语法提示功能不错，适合初学者使用。

现在你可以打开 **IDLE** 新建一个 **py** 为扩展名的 Python 脚本文件，输入以下内容：



保存并运行它，如果输出>>>Hello ,World!，这说明你已经成功编写了第一个 Python 程序，恭喜你！为了比较 Python 与 Java 编码风格等方面的差异，下边给出一个稍微复杂些的”Hello world”程序以及它的 Java 对照版本。

```
1 # -*- coding:utf-8 -*-
2 """
3 我的第一个Python程序
4 """
5 __author__ = 'chuguangming'
6 import sys
7 def Main():
8     sys.stdout.write("Hello world!楚广明\n")
9 if __name__=="__main__":
10     Main()
```

请注意第一行代码是为了支持中文，没有这一行代码的话，我们在下面如果输出中文就会出错的。

Java 的实现

```
/***
 * @author ChuguangMing
 *
 */
public class myfirstjava
{
    public static void main(String[] args)
    {
        System.out.println("这是 Java 的第一个程序");
    }
}
```

Python 的实现

```
 #-*-coding:utf-8-*-
"""
我的第一个应用程序
"""
```

```
import sys
def Main():
    sys.stdout.write("Hello World 我的第一个程序\n")
#下面的语句看起来比较奇怪，一会儿我们会解析它
if __name__=="__main__":
    Main()
```

1. 为什么要学习 Python

我学习过很多语言，C#、Java、C/C++、PHP，但是效高即高而且简单易懂的也就数 Python 了，可以说它是一种胶水语言，我们可以通过它将许多语言整合在一起使用。

我们来看一段代码：

```
-*-coding:utf-8-*-
"""
my fist App
"""

import sys
import urllib
def Main():
    htmlresult=urllib.urlopen("http://www.baidu.com").read()
    print htmlresult
#this is test
if __name__=="__main__":
    Main()
```

2. 代码块与缩进的差异

Java 使用 C/C++风格的编码形式，除了要求用{}组织代码块外，语句间的缩进可以是任意的。

Python 强制所有程序都有相同的编码风格，它通过缩进来组织代码块，缩进相同的语句被认为是处于同一个代码块中，在 if/else 等语句及函数定义式末尾会有一个冒号，指示代码块的开始。Python 这种强制缩进的做法可以省去{}或者 begin/end 等，使程序的结构更为清晰（有的人认为恰好相反），同时也减少了无效的代码行数。

此外需要注意，尽量使用 4 个空格作为 Python 代码的一个缩进单位，最好不要使有 TAB，更不要混用 Tab 和空格，这也算是 Python 的一个非强制性约定吧。

3. 语句结尾的差异

Java 用分号结尾，Python 不用任何符号（类似 BASIC）。实际上 Python 也可以使用分号结尾，像这样 a=1;b=2;c=3;print a,b,c 不过 Python 中这种风格多用于调试，应为你可以很容易注释掉这一行就删除了所有的调试代码。

另外，当一行很长时，Python 可以用\符号折行显示代码。

4. 注释方法

java//用单行注释，用/**/进行多行注释，而 Python 用#符号进行单行注释，用三引号（可单可双）进行多

行注释。

java 的条件表达式必须要加括号，而 **Python** 的条件表达式加不加括号均可。

5. 入口方法

java 语言必须要有入口方法 **Main()**，这是程序开始执行的地方。**Python** 语言中没有入口方法（函数），作为解释型的语言，**Python** 代码会自动从头执行。

如果你对这点不习惯，可以使用 **Python** 代码的内置属性 **__name__** 此属性会根据 **Python** 代码的运行条件变化：当 **Python** 代码以单个文件运行时，**__name__** 便等于 **__main__**，当你以模块形式导入使用 **Python** 代码时，**__name__** 属性便是这个模块的名字。

当然，**Python** 中的 **__name__** 属于并不是为了照顾 **C/C++/C#** 程序员的编程习惯而准备的，它主要目的是用于模块测试。想像一下在 **C#** 中编写一个组件或类代码时，一般还得同时编写一个调用程序来测试它。而 **Python** 中可以把二者合二为一，这就是 **__name__** 属性的真正作用。

6. import 与 using 语句

在用 **Python** 写代码时，我们首先 **import sys**，这是导入了 **Python** 的 **sys** 模块，然后在代码里我们可以引用 **sys** 模块中的对象 **stdout** 及它的 **write** 方法。在 **Python** 中这是必须的，否则你无法调用 **sys** 模块中的任何东西。

简单的说，**Python** 中的 **import** 相当于 **java** 中的包引用。最后 **import** 可以出现在代码的任何位置，只要在引用它之前出现就可以了，不过为了提高程序可读性，建议还是在所有代码开头书写 **import**

2) 小结

- 1、**Python** 使用强制缩进的编码风格，并以此组织代码块
- 2、**Python** 语句结尾不用分号
- 3、**Python** 标明注释用#(单行)或三引号(多行)
- 4、**Python** 语言没有入口方法(**Main**)，代码会从头到尾顺序执行
- 5、**Python** 用 **import** 引用所需要的模块

二、 基本数据类型

1) 第一个程序的解析

“一切数据是对象，一切命名是引用”

如果你能理解这句话，说明对 **Python** 的变量与数据类型已经有了不错的认识，与 **Java** 不同，**Python** 在使用变量之前无须定义它的类型，试着运行下面的例子：

```
#-*-coding:utf-8-*-
"""
我的第一个应用程序
"""

import sys
```

```
def Main():
    sys.stdout.write("开始程序")
    i=1
    print i
#下面的语句看起来比较奇怪，一会儿我们会解析它
if __name__=="__main__":
    Main()
```

从上边我们可以看到，变量 `i` 在使用前并不需要定义，但是必须声明以及初始化该变量。试着运行下面的例子：

```
i=1
print i+j
```

上面的代码会产生一个异常：“`NameError: name 'j' is not defined`”，Python 提示变量 `j` 没有定义。这点和 BASIC 等弱类型的语言不一样。

另一方面，Python 与 java 有一个很大的差异就是在程序运行过程中，同一变量名可以（在不同阶段）代表不同类型的数据，看看下面的例子：

```
-*-coding:utf-8-*-
"""
我的第一个应用程序
"""

import sys
def Main():
    sys.stdout.write("开始程序\n")
    i=1
    print i,type(i),id(i)
    i=1000000000
    print i,type(i),id(i)
    i=1.1
    print i,type(i),id(i)
#下面的语句看起来比较奇怪，一会儿我们会解析它
if __name__=="__main__":
    Main()
```

变量 `i` 的类型在程序执行过程中分别经历了 `int long float` 的变化，这和静态类型语言如 C 语言有很大不同。静态语言只要有一个变量获取了一个数据类型，它就会一直是这个类型，变量名代表的是用来存放数据的内存位置。而 Python 中使用的变量名只是各种数据及对象的引用，用 `id()` 获取的才是存放数据的内存位置，我们输入的 `1`、`1000000000`、`1.1` 三个数据均会保存在 `id()` 所指示的这些内存位置中，直到垃圾回收把它们拉走，这是动态语言的典型特征，它确定一个变量的类型是在给它赋值的时候。

另一方面，Python 又是强类型的，试着运行下边的例子：

```
-*-coding:utf-8-*-
"""
我的第一个应用程序
"""

import sys
def Main():
    sys.stdout.write("开始程序\n")
    i=10
    j='虎虎'
```

```
print i+j
#下面的语句看起来比较奇怪，一会儿我们会解析它
if __name__=="__main__":
    Main()
```

这个程序会出错，产生一个异常”`TypeError:unsupported operand type(s) for +:'int' and 'str'`”。一个正确的写法是：

```
#-*-coding:utf-8-*-
"""
我的第一个应用程序
"""

import sys
def Main():
    sys.stdout.write("开始程序\n")
    i=10
    j='虎虎'
    print str(i)+j
#下面的语句看起来比较奇怪，一会儿我们会解析它
if __name__=="__main__":
    Main()
```

所以，我们说 `Python` 即是一种动态类型语言，同时也是一种强类型的语言，这点是和 `java` 不同的地方，对于 `Python` 的这种变量的声明、定义、使用方式。`Java` 程序员可能要花一段时间去适应，不过相信你会很快喜欢上它，因为它让事情变的更加简单（而且不会不安全）。

我们再小试牛刀一下，比如我们想看看 `c` 盘下面有什么数据与目录的话：

```
#-*-coding:utf-8-*-
import sys
import os
def main():
    sys.stdout.write("Hello World 我的第一个程序\n")
    print os.listdir("c:/")
if __name__=="__main__":
    main()
```

2) 变量的命名规范

`Python` 与 `java` 的变量（函数、类等其他标识符）的命名规范基本一样，同样对大小这与敏感。不一样的地方是 `Python` 中以下划线开始或者结束的标识符通常有特殊的意义。例如以一个下划线开始的标识符如 `_foo` 不能用 `from module import *` 语句导入。前名均有两个下划线的标识符，如 `__init__` 被特殊方法保留。前边有两个下划线的标识符，如 `__bar`，被用来实现类私有属性，这个将在“类和面对对象编程”中再说。

最后，`Python` 的关键字不能作为标识符，不过 `Python` 的关键字比 `java` 要少得多，可以 `google` 一下，这里就列出了。

`Python` 没有常量，如果你非要定义常量，可以引用 `const` 模块

`Python` 程序中一切数据都是对象，包括自定义对象及基本数据类型，这一点和 `C#`一样，它们都是完全面向对象的语言，所以我想 `C#`程序员很容易理解 `Python` 的一切数据都是对象这个口号。

`Python` 不区分值类型和引用类型，你可以把所有的类型都理解为引用类型（当然，它们的实现方式不是一样的，这里只是一个类比）。

Python 内建的数据类型有 20 多种，其中有些不常用到的，有些即将合并。本文将主要介绍空类型、布尔类型、整型、浮点型和字符串、元组、列表、集合、字典等 9 种 Python 内置的数据类型。

在这里，我们将前 4 种称为“简单数据类型”，将后 5 种称为“高级数据类型”，实际上 Python 语言本身没有这种叫法，这样分类是我自己设定的，主要是为了和 java 的相关知识对照方便，希望不要误导大家。

3) 空类型

空类型(`None`)表示该值是一个空对象，比如没有明确规定返回值的函数就是返回 `None`。空类型没有任何属性，经常被用做函数中可选参数的默认值。`None` 的布尔值为假。

Python 的 `None` 和 java 中的可空类型 `Nullable<T>`类似，比如 java 可以定义 `Nullable<double> i=null`，与 Python 的空类型类似，但实现原理和用途都不一样。

4) 布尔类型

Python 中用 `True` 和 `False` 来定义真假，你可以直接用 `a=True` 或 `a=False` 来定义一个布尔型变量，但在 Python2.6 里，`True/False` 以及 `None` 却都不是关键字，在 Python3.0 里它们已经是关键字了，这个有点乱，我们可以不用管它，直接使用就 OK 了。

注意和 java 不同的是，Python 中 True 和 False 的首字母要大写。

最后一点，在 java 中布尔类型和其他类型之间不存在标准的转换。但是在 Python 中，`None`、任何数值类型中的 0、空字符串''，空元组 ()，空列表 []，空字典 {} 都被当作 `False`，其他对象均为 `True`，这点和 C++ 差不多，要提起注意，请思考一下，下面的 Python 代码会输出什么？

```
-*-coding:utf-8-*-
"""
我的第一个应用程序
"""

import sys
def Main():
    sys.stdout.write("开始程序\n")
    if 0:
        print 'True'
    else:
        print 'False'
#下面的语句看起来比较奇怪，一会儿我们会解析它
if __name__=="__main__":
    Main()
```

5) 数值类型

Python 拥有四种数值类型：整型、长整型、浮点类型以及复数类型。

整数类型 (`int`) 用来表示从 -2147483648 到 2147483647 之间的任意整数（在某些电脑系统上这个范围可能会更大，但绝不会比这个更小）；长整型 (`long`) 可以表示任意范围的整数，**实际上我们把 Python 的 long 和 int 理解为同一种类型就可以了，因为当一个整数超过 int 的范围后，Python 会自动将其升级为长整型。**

Python 中只有 64 位双精度浮点数 `float`，与 java 中的 `double` 类型相同（注意在 Python 中浮点数类型名字是 `float` 而不是 `double`），Python 不支持 32 位单精度的浮点数。

除了整数和实数，Python 还提供了一种特殊类型复数(`complex`)。复数使用一对浮点数表示，复数 z 的实部

和虚部分分别用 `z.real` 和 `z.imag` 访问。

在数值运算中，整数与浮点数运算的结果是浮点数，这就是所谓的“提升规则”，也就是“小”类型会被提升为“大”类型参与计算。这一点和 `java` 是一样的，**提升的顺序依次为：int long float complex**

作为数值类型的最后一个问题，`java` 程序员需要注意的是，**Python 没有内建的 decimal 类型**，但可以导入 `decimal` 模块用来完成与货币相关的计算。

`Python` 中序列是由非负整数索引的对象的有序集合，它包括字符串、`Unicode` 字符串，列表、元组、`xrange` 对象以及缓冲区对象。

6) 字符串类型

`Python` 拥有两种字符串类型：标准字符串(`str`)是单字节字符序列，`Unicode` 字符串(`unicode`)是双字节字符序列。

在 `Python` 中定义一个标准字符串(`str`)可以使用单引号、双引号、三引号，这使得 `Python` 输入文本比 `java` 更方便。比如当 `Str` 的内容中包含双引号时，就可以用单引号定义，反之亦然。当字符中有换行符等特殊字符时，可以直接使用三引号定义。这样就方便了很多不用去记那些乱七八糟的转义字符。当然 `Python` 也支持转义字符，且含义与 `java` 基本一样。

下面是一个例子来说明这一点

```
#-*-coding:utf-8-*-
"""
我的第一个应用程序
"""

import sys
def Main():
    sys.stdout.write("开始程序\n")
    str1='i am "python"\n'
    str2="i am 'Python' \r"
    str3=""""
        i'm "Python",
        <a href="http://www.sina.com.cn"></a>
    """
    #你可以把 html 之类的东西都直接弄进来而不需要处理
    print str1,str2,str3
if __name__=="__main__":
    Main()
```

在 `Python` 中定义一个 `Unicode` 字符串，需要在引号前面加上一个字符 `u`，例如

```
#-*-coding:utf-8-*-
import sys
def Main():
    print u"我是派森"
if __name__=="__main__":
    Main()
```

同时注意，当使用 `UTF-8` 编码时，非 `unicode` 字符中一个汉字的长度是 3，而使用 `gb2312` 时是 2，见下边代码：

```
#-*-coding:utf-8-*-
...
test
...
import sys
```

```
def Main():
    print u"我是派森"
    unicode =u'我'
    str='我'
    print len(unicode),len(str)
    #输出的 1 3
if __name__=="__main__":
    Main()
```

7) 全局变量

不加入 `global` 关键字的话，如果你是在函数内部定义的变量的话，在外部是无法获取的。

```
#coding:utf-8
def testList():
    global list
    list=[]
    list.append("test1")
    list.append("test2")
    print list
testList()
print list
```

三、 基本数据类型-列表(list []))

1) 概述

Python 中列表 (`list`) 类似于 java 中的 `ArrayList`，用于顺序存储结构。列表用符号`[]`表示，中间的元素可以是任何类型（包括列表本身，以实现多维数组），元素之间用逗号分隔。取值或赋值的时候可以像 C 数组一样，按位置索引：

```
-*-coding:utf-8-*-
import sys
def Main():
    array=[1,2,3]
    print array[0]
    #输出 1
    array[0]='a'
    print array
    #输出['a',2,3]
    L=[123,'spam',1.23]
    #输出大小
    print len(L)
    print L[0]
    print L[:-1]#不包含最后一个
```

```
print L+[4,5,6]#重新拼接一个新的列表
if __name__=="__main__":
    Main()
```

从上边的代码中你可能发现一个有趣的事情：在 Python 的列表中可以混合使用不同类型的数据，像 [‘a’, 2, 3] 这样，不过我不建议你这样做，我觉着没有什么好处。

另外还可以看到，列表是可变的序列，也就是说我们可以在“原地”改变列表上某个位置所存储的对象的值。

Python 中的 list 支持多数的操作，同时 list 也支持“切片”这样的操作。切片指的是抽取序列的一部分，其形式为：list[start:end:step]。其抽取规则是：从 start 开始，每次加上 step，直到 end 为止。默认的 step 为 1；当 start 没有给出时，默认从 list 的第一个元素开始；当 end=-1 时表示 list 的最后一个元素，依次类推。一些简单的例子见下边代码：

```
-*-coding:utf-8 -*-
import sys
def Main():
    test=['never',1,2,'yes',1,'no','maybe']
    print test[0:3]#包括 test[0],不包括 test[3]
    print test[0:6:2]#包括 test[0],不包括 test[6],而且步长为 2
    print test[::-1]#包括开始, 不包括最后一个
    print test[-3:]#抽取最后 3 个
if __name__=="__main__":
    Main()
```

字符串、列表、元组都支持切片操作，这个很方便，应该学会熟练使用它。

最后，list 是 Python 中最基础的数据结构，你可以把它当作链表、堆栈或队列来使用，效率还不错。Python 中没有固定长度数组，如果你确实很在意性能，可以改入 array 模块来创建一个 C 风格的数组，它的效率很高，这里就不详细介绍了。

我们还可以对其进行排序与反转

```
-*-coding:utf-8-
import sys
def Main():
    array=[5,2,3,1,8]
    array.sort()
    for s in array:
        print s
    array.reverse()
    for s in array:
        print s
if __name__=="__main__":
    Main()
```

Python 核心数据类型的一个优秀的特性就是它们支持任意的嵌套。能够以任意的组合对其进行嵌套。这种特性的直接应用就是实现矩阵，或者 Python 中的多维数组。一个嵌套列表的矩阵能够完成这个基本的操作。

```
-*-coding:utf-8-
import sys
def Main():
    M=[[1,2,3],
       [4,5,6],
       [7,8,9]]
    print M[0]
```

```
print M[1]
print M[2]
if __name__=="__main__":
    Main()
```

处理序列的操作和列表的方法中，Python 还包括了一个更高级的操作，称作列表解析表达式，从而提供了一种处理像矩阵这样结构的强大工具。例如，假设我们需要从列举的矩阵中提取出第二列。因为矩阵是按照行进行存储的，所以可以通过简单的索引即可获得行，使用列表解析可以同样简单地获得列。

```
-*-coding:utf-8 -*-
import sys
def Main():
    M=[[1,2,3],
       [4,5,6],
       [7,8,9]]
    col2=[row[1] for row in M]
    print col2
    col3=[row[1]+1 for row in M]
    print col3
    colfilter=[row[1] for row in M if row[1]%2==0]
    print colfilter
if __name__=="__main__":
    Main()
```

2) 常用方法

`len(list)` 长度
`del list` 删除对象

列表对象支持的方法：

`append(x)` 尾部追加 单个对象 x，使用多个对象会引起异常。
`count(x)` 返回对象 x 在 list 中出现的次数
`extend(L)` 将列表 L 中的项添加到表中
`index(x)` 返回匹配对象 x 第一个表项的索引，无匹配时产生异常
`insert(i,x)` 在索引 ‘i’ 的元素前插入对象 x
`pop(x)` 删除列表中索引 x 的表项，并返回同该表项的值，无参数删除最后
`remove(x)` 删除表匹配对象 x 的第一个元素，无匹配时异常
`reverse()` 颠倒列表元素的顺序
`sort()` 对列表排序

此外可简单使用+实现列表连接： `[3,4] + [[1,2],5,6] --> [3,4,[1,2],5,6]`

删除列表中的重复项：`M = list(set(L))`，python 的 set 和其他语言类似，是一个无序不重复元素集

四、 基本数据类型-字典(dict { })

1) 概述

用过 java 中的字典的人对 Hashtable 应该不会陌生，Python 里的哈希表就是字典(dict)了。与 set 类似，字典是一种无序存储结构，它包括关键字(key)和关键字对应的值(value)。

java 程序员需要了解的就是，在 Python 中 dict 是一种内置的数据类型，定义方式为：

```
dictionary={key:value}#当有多个键值时，用逗号进行分割。
```

字典里的关键字为不可变类型，如字符串、整数、只包含不可变对象的元组，列表等不能作为关键字。字典中一个键只能与一个值关联，对于同一个键，后添加的值会覆盖之前的值。

学过数据结构的人对字典的散列查找效率应该都有认识，所以我建议在可能的情况下尽量多用字典，其它的就不多写了。

```
-*-coding:utf-8-*-
import sys

if __name__=="__main__":
    dict={"a":"apple","b":"banana","g":"grape","o":"orange"}
    print dict
    print dict["a"]
    dict2={1:"apple",2:"banana",3:"grape",4:"orange"}
    print dict2
    print dict2[1]
```

```
-*-coding:utf-8-*-
import sys

if __name__=="__main__":
    #字典的添加、删除、修改操作
    dict={"a":"apple","b":"banana","g":"grape","o":"orange"}
    dict["w"]="watermelon"
    print dict
    del(dict["a"])
    print dict
    print dict.pop("b")
    #dict.clear()
    #print dict
    #字典的遍历
    for k in dict:
        print "dict[%s]=\"%s",dict[k]
```

```
-*-coding:utf-8-*-
import sys

if __name__=="__main__":
```

```
#字典的 keys() 与 values() 方法
dict={"a":"apple","b":"banana","g":"grape","o":"orange"}
#输出 key 的列表
print dict.keys()
#输出 values 的列表
print dict.values()
```

```
-*-coding:utf-8 -*-
import sys
def Main():
    D={'food':'spam','quantity':4,'color':'pink'}
    print D['food']
    D['quantity']+=1
    print D
    #另外一种定义字典的方法
    D={}
    D['name']='Bob'
    D['job']='dev'
    D['age']=40
    print D
    #使用键值,进行排序
    D={'a':1,'b':2,'c':3}
    print D
    Ks=D.keys()
    print Ks
    Ks.sort()
    print Ks
    for key in Ks:
        print key,'=>',D[key]
    for key in sorted(D):
        print key,'=>',D[key]
    #迭代与优化
    squares=[x ** 2 for x in [1,2,3,4,5]]
    print squares
    #与以下代码是等效的
    squares=[]
    for x in [1,2,3,4,5]:
        squares.append(x**2)

if __name__=="__main__":
    Main()
```

在访问的时候，如果这个键值不存在的话，如果我们没有做任何判断的话，会出现错误，这个时候我们可以用以下代码来进行判断。

```
-*-coding:utf-8-
import sys
```

```
def Main():
    D={'food':'spam','quantity':4,'color':'pink'}
    #测试不存在的键值
    if not D.has_key('f'):
        print '不存在这个键值'
    else:
        print D['f']

if __name__=="__main__":
    Main()
```

2) 常用方法

```
has_keys(x) 若字典中有 x 返回 true
keys() 返回键的列表
values() 返回值的列表
dict.items() 返回 tuples 的列表。每个 tuple 有字典的 dict 的键和相应的值组成
clear() 删除词典的所有条目
copy() 返回字典的高层结构的拷贝，但不复制嵌入结构，而复制那些结构的引用。
update(x) 用字典 x 中的键/值对更新字典的内容。
get(x[,y]) 返回键 x。若未找到返回 None
```

五、 基本数据类型-元组(tuple())

元组与列表非常相似，它用()而不是[]括起来的序列。元组比列表的速度更快，但元组是一个不可变的序列，也就是与 str 一样，无法在原位改变它的值。除此之外，其他属性与列表基本一致。

元组是 Python 中内置的一种数据结构。元组由不同的元素组成，每个元素可以存储不同类型的数据，如字符串、数字甚至元组。元组是写保护的！即元组创建后不能再做任何修改操作，元组通常代表一行数据，而元组中的元素代表不同的数据项。

元组的创建

```
-*-coding:utf-8-*-
import sys
if __name__=="__main__":
    tuple_name=("apple","banana","grape","orange")
    print tuple_name[0]
```

分片输出

```
-*-coding:utf-8-*-
import sys
if __name__=="__main__":
    tuple_name=("apple","banana","grape","orange")
    print tuple_name[1]
```

```
#牛在支持分片输出
print tuple_name[-1]
print tuple_name[-2]
print tuple_name[1:3]
#我还可以在元组中包含自己
print '-----'
tuple=([('t1','t2'),('t3','t4'))
print tuple[0][0]
print tuple[1][0]
```

实现解包的功能

```
-*-coding:utf-8-*
import sys
if __name__=="__main__":
    tuple_name=("apple","banana","grape","orange")
    a,b,c,d=tuple_name
    print a,b,c,d
```

元组定义的方法与列表类似，不过在定义只包含一个元素的元组时，注意在后边加一个逗号，请体会以下几句的差异：

```
-*-coding:utf-8-*
import sys
def Main():
    test=[0]          #列表可以这样定义
    print type(test) #输出<type 'list'>
    test=[0,]         #也可以这样定义
    print type(test) #输出<type 'list'>
    test=(0,)         #元组可以这样定义
    print type(test) #输出<type 'tuple'>
    test=(0)          #但不能这样定义，Python 会认为它是一个括号表达式
    print type(test) #输出<type 'int'>
    test=0,           #也可以省略括号，但要注意与 C 的逗号表达式不同
    print type(test) #输出<type 'tuple'>
#还可以简单的交换数据
a=1
b=2
a,b=b,a
print a,b
if __name__=="__main__":
    Main()
```

以上这类语句在 Python 中被广泛应用于变量交换、函数传值等应用，因此 Python 的解析器在不断对其进行优化，现在已经具备了相当高的效率。所以上代码在 Python2.5 以后的版本中，比 `tmp=a;a=b;b=tmp` 这种常规语句更快。

- Python 是一种动态的强类型语言，在使用变量之前无须定义其类型，但是必须声明和初始化
- “一切命名是引用”，Python 中变量名是对象的引用，同一变量名可以在程序运行的不同阶段代表不同类型的数据
- “一切数据都是对象”，Python 的所有数据类型都是对象，相较 java 具有一致的使用方法

- 把问题想得更简单一点，Python 的数值类型可以说只有两种：整形和浮点
- 多使用 `list/tuple/set/dict` 这几种很 `pythonic` 的数据类型，它们分别用`[]/()/([])/{}`定义

六、 基本数据类型-集合(set)

Python 中的 `set` 和 java 中的集合不是一个概念，这是翻译的问题，Python 中的集合是指无序的、不重复的元素集，类似数学中的集合概念，可对其进行交、并、差、补等逻辑运算。

常见集合的语法为：`s=set(['a','b','c'])`。不过 `set` 在 Python3.0 中发生了较大的变化，创建一个集合的语法变成了：`S={1,2,3}`，用花括号的方法，与后边要提到的 `dict` 类似。

如果在 `set` 中传入重复元素，集合会自动将其合并。这个特性非常有用，比如去除列表里大量的重复的元素，用 `set` 解决效率很高。示例如下：

```
#-*-coding:utf-8-*-
import sys
def Main():
    a=[133,224,2344,2243,22342,224,133,133,989]
    b=set(a)
    print b
if __name__=="__main__":
    Main()
```

另一个例子，找出两个 `list` 里面相同的元素(集合求交，其它类推)，代码如下：

```
#-*-coding:utf-8-*-
import sys
def Main():
    a=[133,224,2344,2243,22342,224,133,133,989]
    b=set(a)
    print b
    a=["11","22","33"]
    b=["11","33"]
    c=set(a)&set(b)
    print c
if __name__=="__main__":
    Main()
```

想想你如果自己实现这个算法会怎么写？然后可以找两个大一点的列表，比比和 `set` 实现的效率，你就会有体会了，在以后的程序中多用 `set` 吧。

七、 数据类型转换

1) 概述

2) 常用方法

```
str(x) 将对象 x 翻译为字符串  
list(x) 将对象序列 x 作为列表返回。例如 ‘hello’ 返回['h','e','l','l','o']，将 tuple 转换为列表  
tuple(x) 将对象序列 x 作为 tuple 返回  
int(x) 将字符串和数字转换为整数，对浮点进行舍位而非舍入  
long(x) 将字符串和数字转换为长整形  
float(x) 将 str 和 num 转换为浮点对象  
complex(x,y) 将 x 做实部，y 做虚部创建复数  
hex(x) 将整数或长整数转换为十六进制字符串  
oct(x) 将整数或长整数转换为八进制字符串  
ord(x) 返回字符 x 的 ASCII 值  
chr(x) 返回 ASCII 码 x 代表的字符  
min(x[,...]) 返回序列中最小的元素  
max(x[,...]) 返回序列中最大的元素
```

八、 运算符、表达式和流程控制

本章介绍 Python 的运算符、表达式、程序流程控制语句以及异常处理语句，在这方面，Python 和 java 是非常类似的，我们仅需要注意它们之间的一些细微差异。另外，在本章我还会简要介绍 Python 语言中的两项有趣功能=列表内涵和动态表达式，虽然它们严格来说属于函数部分的内容，不过我觉得还是放在表达式一章比较合适。

无论使用什么语言，我们编写的大多数代码都包含表达式。一个表达式可以分解为运算符和操作数，运算符的功能是完成某件事，它们由一些数学运算符号或者其他特定的关键字表示，运算符需要数据来进行运算，这样的数据被称为操作数。例如，`2+3` 是一个简单的表达式，其中`+`是运算符，`2` 和 `3` 是操作数。

1) 算术运算符与算术表达式

算术运算符是程序设计语言最基本的运算符。Python 提供的算术运算符除了`+`、`-`、`*`、`/`、`%`（求余）之外，还提供了两种 java 中没有提供的运算符：求幂(`**`)和取整除(`//`)。下面我们就通过一段代码来解析这两个算术运算符的功能。

```
-*-coding:utf-8-*  
import sys  
def Main():  
    x=3.3  
    y=2.2  
    a=x**y  
    print a  
    #输出即 3.3 的 2.2 次幂  
    b=x//y  
    print b  
    #输出 1.0 取整除返回商的整数部分  
    c=x/y
```

```
print c
#输出 1.5, 注意体会普通除与取整除的区别
if __name__=="__main__":
    Main()
```

2) 赋值运算符与赋值表达式

赋值就是给一个变量赋一个新值，除了简单的=赋值之外，Python 和 java 都支持复合赋值，例如 `x+=5`, 等价于 `x=x+5`.

Python 不支持 java 中的自增和自减运算符，例如 `X++` 这种语句在 Python 中会被提示语法有错误。java 程序员可能习惯了这种表达式，在 Python 中，请老老实实的写 `X+=1` 就是了。

Python 的逻辑运算符与 java 有较大区别，Python 用关键字 `and or not` 代替了 java 语言中的逻辑运算符 `&& || !`，此外 Python 中参与逻辑运算符的操作数不限于布尔类型，任何类型的值都可以参与逻辑运算中去。

用逻辑运算符将操作数或表达式连接起来就是逻辑表达式。与 java 一样，Python 中逻辑表达式是短路执行的，也就是说只有需要时才会进行逻辑表达式右边值的计算，例如表达式 `a and b` 只有当 `a` 为 `true` 时才计算 `b`. 思考一下，`if(0 and 10/0)`: 这条语句会引发除数为零的异常吗？

此外还要注意：在 Python 中，`and` 和 `or` 所执行的逻辑运算并不返回布尔值，而是返回它们实际进行比较的值之一。下边是一个例子：

```
 #-*-coding:utf-8-*-
import sys
def Main():
    print 'a' and 'b'
    print '' and 'b'
if __name__=="__main__":
    Main()
```

3) 关系运算符与关系表达式

关系运算符实际上是逻辑运算的一种，关系表达式的返回值总是布尔值。Python 中的比较操作符与 java 完全是一样的，包括 `== != > < >= <=` 共 6 种。

除了基本的变量比较外，Python 的关系运算符还包括身份运算符 `is`. 在 Python 中，`is` 用来检验两个对象在内存中是否指向同一个对象（还记得一切数据皆对象吗？一切命名皆引用吗）。

三元运算符

三元运算符是 `c/c++` 系语言所特有的一类运算符，例如，对表达式 `b?x:y`，先计算条件 `b`，然后进行判断，如果 `b` 的值为 `true`，则计算并返回 `x` 的值，否则计算并返回 `y` 的值。

在 Python 中，提供了专门的逻辑分支表达式来模拟 java 系中的三元运算，我们也可以在一行语句中完成三元运算，例如

```
print '偶数' if x%2==0 else '奇数'
```

九、 流程控制

1) 条件语句

Python 用 `if`, `elif`, `else` 这三个关键字进行条件判断，与 Java 唯一的区别就是用 `elif` 取代 `else if`，少打两个字，其它都一样，此外别忘了在 `if` 等语句后加：号

如果一个流程控制分南下不做任何事情，记得写一句 `pass` 语句，不然 Python 会报错。例如：

```
if 0:  
    pass #这一句语没有什么意义
```

在 Python 中没有 `switch` 语句，你可以使用 `if..elif..else` 语句来完成同样的工作。如果你觉得繁琐，可以试试 `dict` 实现的方式，下边是一个例子，分别对比了两种实现方式。

```
-*-coding:utf-8-*  
import sys  
def Main():  
    #使用 if 替代  
    x='4'  
    print "OK"  
    if x=='1':  
        print 'one'  
    elif x=='2':  
        print 'two'  
    else:  
        print 'nothing!'  
    #使用 dict  
    numtrans={  
        1:'one',  
        2:'two',  
        3:'three'  
    }  
    try:  
        print numtrans[x]  
    except KeyError:  
        print 'nothing!'  
if __name__=="__main__":  
    Main()
```

2) 循环

Python 支持两种循环语句 `while` 循环和 `for` 循环，不支持 Java 中的 `do-while` 循环。在 Python 的 `while` 循环和 Java 基本一致，此处我们着重比较两种语言中的 `for` 循环的区别。

说的简单一点，python 中的 `for` 语句相当于 java 中的 `foreach` 语句，它用于从集合对象（list/str/tuple 等）中遍历数据。例如：

```
for I in [1,2,3,4,5]:  
    print i
```

```
for i in range(10):
    print i
```

```
-*-coding:utf-8 -*-
import sys
if __name__=="__main__":
    tuple=(("apple","banana"),("grape","orange"))
    for i in range(50,100+1):
        print i
```

```
-*-coding:utf-8-
import sys
if __name__=="__main__":
    for i in range(50,100+1):
        print i
range(1,5) #代表从 1 到 5(不包含 5)
range(1,5,2) #代表从 1 到 5, 间隔 2(不包含 5)
range(5) #代表从 0 到 5(不包含 5)
```

```
__author__ = 'TenYear'
-*-coding:utf-8-*
"""
my fist App
"""

import sys
import urllib
def Main():
    itemlist=[1,2,3,4,5,4]
    for m in itemlist:
        print m
#this is test
if __name__=="__main__":
    Main()
```

十、 异常

Python 和 java 一样支持异常处理，利用 `try/except/finally` 结构，可以很方便的捕获异常，同时可以用 `raise` 语句手动抛出异常（上述四个异常处理的关键字分别对应 Java 中的 `try/catch/finally/throw`）。通过 `except`，您可以将 `try` 标示的语句中出现的错误和异常捕获。

```
-*-coding:utf-8-
import sys
def Main():
    try:
        f=open('firstpython.py')
```

```
s=f.readline()
print s
except IOError,(errno,strerror):
    print "I/O error(%s):%s" %(errno,strerror)
except ValueError:
    print "Could not convert data to an integer"
except:
    print "Unexpected error:",sys.exc_info()[0]
    raise
finally:
    f.close()
if __name__=="__main__":
    Main()
```

最后说明一点，`python` 的 `try` 也支持 `else` 语句，如果有一些代码要在 `try` 没有发生异常的情况下才执行，就可以把它放在 `else` 中。

十一、 动态表达式

在 C# 语言中，如果需要在文本框中输入 `1+2`（或更加复杂的数学表达式）后计算它的值，可能会用到表达式解析等。现在我们有了 `Python`，要完成这种任务可以说是非常简单：只要用内置的 `eval()` 函数，就可以计算并返回任意有效表达式的值。例如：

```
str='1+2'
print eval(str)
```

除了 `eval` 函数之外，`Python` 还提供了 `exec` 语句将字符串 `str` 当成有效 `Python` 代码来执行，看下面的例子：

```
exec 'a=100'
print a
```

另外还有 `execfile` 函数，它用来执行一个外部的 `py` 文件，上一个例子存为 `exec.py` 后，运行下边的代码就知道是怎么回事了：

```
execfile(r'c:\exec.py')
```

最后提醒，默认的 `eval()`, `exec`, `execfile()` 所运行的代码都位于当前的名字空间中，`eval()`, `exec`, 和 `execfile()` 函数也可以接受一个或两个可选字典参数作为代码执行的全局名字空间和局部名字空间，具体可以参考 `Python` 手册。

列表内涵是 `Python` 最强有力的语言之一，常用于从集合对象中有选择地获取并计算元素，虽然多数情况下可以使用 `for/if` 等语句组合完成同样的任务，但列表内涵书写的代码更加简洁。

```
 #-*-coding:utf-8-*-
import sys
def func():
    x=1
    y=2
    m=3
    n=4
    sum=lambda x,y:x+y
    print sum
    sub=lambda m,n:m-n
```

```
print sub
return sum(x,y)*sub(m,n)
if __name__=="__main__":
    print func()
```

列表内涵的一般形式如下，我们可以把[]内的列表内涵写为一行，也可以写为多行。

[表达式 for item1 in 序列1 ... for itemN in 序列N if 条件表达式]

上面的表达式分为三部分，最左边是生成每个元素的表达式，然后是 for 迭代过程，最右边可以设定一个 if 判断作为过滤条件。

列表内涵的一个著名例子是生成九九乘法表：

```
s=[(x,y,x*y) for x in range(1,10) for y in range(1,10) if x>=y]
```

十二、 函数及函数编程

在 java 中没有独立的函数存在，只有类的（动态或静态）方法这一概念，它指的是类中用于执行计算或其它行为的成员。在 Python 中，你可以使用类似 C# 的方式定义类的动态或静态成员方法，因为它与 java 一样支持完全的面向对象编程。你也可以用过程式编程的方式来编写 Python 程序，这时 Python 中的函数与类可以没有任何关系，类似 C 语言定义和使用函数的方式。此外，Python 还支持函数式编程，虽然它对函数编程的支持不如 LISP 等语言那样完备，但适合使用还是可以提高我们工作的效率的。

1) 函数的定义

函数定义是最基本的行为抽象代码，也是软件复用最初级的方式。Python 中函数的定义语句由 def 关键字、函数名、括号、参数及冒号组成。下面是几个简单的函数定义语句：

```
-*-coding:utf-8-*-
import sys
def Main():
    def F1():
        print "我是 F1"
    def F2(x,y):
        a=x+y
        return a
    #定义有多个返回值的函数，用逗号分割不同的返回值，
    #返回结果是一个元组
    def F3(x,y):
        a=x/y
        b=x%y
        return a,b
if __name__=="__main__":
    Main()
```

可能你已经注意到了，Python 定义函数的时候并没有约束参数的类型，它以最简单的形式支持了泛型编程。你可以输入任意类型的数据作为参数，只要这些类型支持函数内部的操作

2) 定义一个交换函数

```
#-*-coding:utf-8-*-
import sys
import random
def compareNum(num1,num2):
    if(num1>num2):
        return 1
    elif(num1==num2):
        return 0
    else:
        return -1
if __name__=="__main__":
    num1=random.randrange(1,9)
    num2=random.randrange(1,9)
    print "num1=",num1
    print "num2=",num2
    print compareNum(num1,num2)
```

3) 函数的默认参数与返回值

```
#-*-coding:utf-8-*-
import sys
def arithmetic(x=1,y=1,operator="+"):
    result={
        "+":x+y,
        "-":x-y,
        "*":x*y,
        "/":x/y
    }
    return result.get(operator)
if __name__=="__main__":
    print arithmetic(1,2)
    print arithmetic(1,2,"-")
```

4) 返回多个值

Python 支持多个返回值

```
#coding:utf-8
def testList():
    return "aaa","bbb"
a,b=testList()
print a
print b
```

5) locals 函数和 globals 函数介绍

Python 有两个内置的函数，`locals()` 和 `globals()`，它们提供了基于字典的访问局部和全局变量的方式。首先，是关于名字空间的一个名词解释。是枯燥，但是很重要，所以要耐心些。Python 使用叫做名字空间的东西来记录变量的轨迹。名字空间只是一个字典，它的键字就是变量名，字典的值就是那些变量的值。实际上，名字空间可以象 Python 的字典一样进行访问，一会我们就会看到。

在一个 Python 程序中的任何一个地方，都存在几个可用的名字空间。每个函数都有着自己的名字空间，叫做局部名字空间，它记录了函数的变量，包括 函数的参数和局部定义的变量。每个模块拥有它自己的名字空间，叫做全局名字空间，它记录了模块的变量，包括函数、类、其它导入的模块、模块级的变量和常量。还有就是内置名字空间，任何模块均可访问它，它存放着内置的函数和异常。

当一行代码要使用变量 `x` 的值时，Python 会到所有可用的名字空间去查找变量，按照如下顺序：

1. 局部名字空间 - 特指当前函数或类的方法。如果函数定义了一个局部变量 `x`，Python 将使用这个变量，然后停止搜索。
2. 全局名字空间 - 特指当前的模块。如果模块定义了一个名为 `x` 的变量，函数或类，Python 将使用这个变量然后停止搜索。
3. 内置名字空间 - 对每个模块都是全局的。作为最后的尝试，Python 将假设 `x` 是内置函数或变量。

如果 Python 在这些名字空间找不到 `x`，它将放弃查找并引发一个 `NameError` 的异常，同时传递 `There is no variable named 'x'` 这样一条信息，象 Python 中的许多事情一样，名字空间在运行时直接可以访问。特别地，局部名字空间可以通过内置的 `locals` 函数来访问。全局（模块级别）名字空间可以通过 `globals` 函数来访问

```
#coding:utf-8
from __future__ import unicode_literals
from sys import *
def test(arg):
    #函数 test 在它的局部名字空间中有两个变量:arg(它的值被传入函数)
    # 和 z(它在函数内部)
    z=1
    print locals()
#locals 返回一个名字与值的字典，这个字典的键字是字符串形式的变量名
#字典的值是变量的实际值
test(5)
test('楚广明')
print globals()#返回一个全局的字典，包括所有导入的变量
```

十三、 类及面向对象

如果你熟悉 `java` 或者 `c#`，那么对类和面向对象应该不会陌生。Python 与 `c#`一样，能够很好的支持面向对象的编程模式。

1) 类的定义

与 C#一样，Python 使用 `class` 关键字定义一个类。一个最简单的类定义语句如下：

```
Class A:  
    Pass
```

它等价于 C#中的 `class A{}`.当然以上语句没有任何实际意义，它只是告诉我们什么是定义一个类所必需的。即 `class` 关键字，类名和冒号，**pass 关键字只用来占位，相当于用来占位，相当于 C#中花括号的作用。**

类是定义对象格式的模板，而对象则是类的实例，通过类创建对象的过程称为类的实例化。在 C#中，需要使用 `new` 关键字实例化一个类，例如

```
A a=new A();
```

在上条语句，C#完成了两件事，首先声明一个类型为 A 的变量 `a`，然后用 `new` 运算符创建一个类型为 A 的对象，并将该对象的引用赋值给变量 `a`，而在 python 中没有 `new` 关键字，同时它是一个动态语言，不需要事先指定变量的类型，只需要：

```
a=A()
```

即创建一个类型为 A 的对象，看起来好像是将类当作一个函数调用，返回值是新创建的对象。

2) 为类添加数据

通常我们利用类来定义各种新的数据类型，其中即包含数据内容，也包含对数据内容的操作。Python 类的数据添加方法与 C#有一些不同，因为 Python 是一种动态语言，变量在使用之前不需要定义，所以你可以不在类定义中添加成员变量，而是在运行时动态地添加它们，例如：

```
-*-coding:utf-8-*  
import sys  
def Main():  
    class A:pass  
    a=A()  
    a.x=1  
    print a.x  
if __name__=="__main__":  
    Main()
```

3) 构造函数

Python 的类提供了类似 C#构造函数的东西：`__init__`（注意是前后是两个下划线），类在实例化时会首先调用这个函数，我们可以通过重写`__init__`函数，完成变量的初始化等工作。与 C#不同的地方是，**Python 不支持无参数的初始化函数，你至少需要为初始化函数指定一个参数，即对象实例本身(self)**。下面是一段简单的代码。在该代码中，我们重写了函数`__init__`，定义并初始化一个类的成员变量 X:

```
-*-coding:utf-8-*  
import sys  
def Main():  
    class A:  
        def __init__(self):  
            self.x=1  
    a=A()  
    print a.x
```

```
if __name__=="__main__":
    Main()
```

```
#-*-coding:utf-8 -*-
import sys
class Student:
    def __init__(self,name,age):
        self.__name=name
        self.__age=age
    def getName(self):
        format="my name is %s my age is %d"%(self.__name,self.__age)
        print format
    def __del__(self):
        print "del"
if __name__=="__main__":
    student=Student("chu",35)
    student.getName()
```

4) 静态成员与私有成员

```
#-*-coding:utf-8 -*-
import sys
class A:
    #定义静态成员变量
    y=2
    def __init__(self):
        #定义公共的成员变量
        self.x=1
        #定义私有的成员变量
        self.__z=1
if __name__=="__main__":
    a=A()
    #打印公共的成员变量
    print a.x
    #打印静态成员变量
    print A.y
    #打印私有成员变量出错
    #print a.__z
```

5) 为类添加方法

方法是对类数据内容的操作，在 Python 中，定义类的方法与定义一个普通的函数在语法上基本上相同，c#程序员需要注意的是，在类中定义的常规方法的第一个参数总是该类的实例，即 `self`，同时注意在方法中引用类的另一个方法必须使用类名加方法名的形式，下面是一个定义类的方法的简单的例子。

```
#-*-coding:utf-8-
```

```
import sys
class A:
    def prt(self):
        print "my name is A"
    def reprt(self):
        A.prt(self)
if __name__=="__main__":
    a=A()
    a.prt()
    a.reprt()
```

不能定义一个不操作实例的方法

```
 #-*-coding:utf-8-*-
import sys
class Student:
    __name=""
    def __init__(self,name):
        self.__name=name
    def getName(self):
        return self.__name
if __name__=="__main__":
    student=Student("chu")
    print student.getName()
```

6) 静态方法

Python 与 C#一样支持静态方法。在 C#中需要使用关键字 `static` 声明一个静态方法，而在 Python 中是通过静态方法修饰符`@staticmethod`来实现的，下面是例子：

```
 #-*-coding:utf-8-*-
import sys
class A:
    def prt(self):
        print "my name is A"
    def reprt(self):
        A.prt(self)
    @staticmethod
    def prt2():
        print "我是静态方法"
if __name__=="__main__":
    a=A()
    a.prt()
    a.reprt()
    A.prt2()
```

如你所见，静态方法可以直接被类调用，它没有常规方法那样的特殊行为（默认第一行参数是 `self` 等），你完全可以将静态方法当成一个用属性引用方式调用的普通函数。

7) 单继承

Python 用类名后加括号的方式实现继承，下面是一个简单的示例：

```
-*-coding:utf-8-*-
import sys
class A:
    x=1
class B(A):
    y=2
if __name__=="__main__":
    print B.x
    print B.y
```

一个复杂的实例

```
-*-coding:utf-8-*-
import sys
class SchoolMember:
    '''Represents any school member.'''
    def __init__(self, name, age):
        self.name = name
        self.age = age
        print'(Initialized SchoolMember: %s)'% self.name

    def tell(self):
        '''Tell my details.'''
        print'Name:"%s" Age:"%s"'% (self.name, self.age),

class Teacher(SchoolMember):
    '''Represents a teacher.'''
    def __init__(self, name, age, salary):
        SchoolMember.__init__(self, name, age)
        self.salary = salary
        print'(Initialized Teacher: %s)'% self.name

    def tell(self):
        SchoolMember.tell(self)
        print'Salary: "%d"'% self.salary

class Student(SchoolMember):
    '''Represents a student.'''
    def __init__(self, name, age, marks):
        SchoolMember.__init__(self, name, age)
        self.marks = marks
        print'(Initialized Student: %s)'% self.name

    def tell(self):
```

```
SchoolMember.tell(self)
print'Marks: "%d"'% self.marks

if __name__=="__main__":
    t = Teacher('Mrs. Shrividya',40,30000)
    s = Student('Swaroop',22,75)
```

8) 传值与传引用

1. Python 不允许程序员选择采用传值还是传引用。Python 参数传递采用的肯定是“传对象引用”的方式。实际上，这种方式相当于传值和传引用的一种综合。如果函数收到的是一个可变对象（比如字典或者列表）的引用，就能修改对象的原始值——相当于通过“传引用”来传递对象。如果函数收到的是一个不可变对象（比如数字、字符串或者元组）的引用，就不能直接修改原始对象——相当于通过“传值”来传递对象。

2. 当人们复制列表或字典时，就复制了对象列表的引用同，如果改变引用的值，则修改了原始的参数。
3. 为了简化内存管理，Python 通过引用计数机制实现自动垃圾回收功能，Python 中的每个对象都有一个引用计数，用来计数该对象在不同场所分别被引用了多少次。每当引用一次 Python 对象，相应的引用计数就增 1，每当消毁一次 Python 对象，则相应的引用就减 1，只有当引用计数为零时，才真正从内存中删除 Python 对象。

```
 #-*-coding:utf-8-*-
def add_list(p):
    p=p+[1]
def add_list2(p2):
    p2+=[1]
tmpList=[1,2,3,4]
add_list(tmpList)
print tmpList
#结果[1,2,3,4]
add_list2(tmpList)
print tmpList
#结果[1,2,3,4,1]
```

这主要是由于“=”操作符会新建一个新的变量保存赋值结果，然后再把引用名指向“=”左边，即修改了原来的 p 引用，使 p 成为指向新赋值变量的引用。而+=不会，直接修改了原来 p 引用的内容，事实上+=和=在 python 内部使用了不同的实现函数。

```
 #-*-coding:utf-8-*-
a=[]
b={'num':0,'sqrt':0}
resurse=[1,2,3]
for i in resurse:
    b['num']=i
    b['sqrt']=i*i
    a.append(b)
print a
#结果[{'num': 1, 'sqrt': 1}, {'num': 2, 'sqrt': 4}, {'num': 3, 'sqrt': 9}]
#这是由于 a 中的元素就是 b 的引用。可以修改为:
```

```
#修改
a=[]
for i in resurse:
    a.append({'num':i,"sqrt":i*i})
print a
```

十四、 模块与包

Python 的脚本都是用扩展名为 **py** 的文本文件保存的，一个脚本可以单独运行，也可以导入另一个脚本中运行。当脚本被导入运行时，我们将其称为模块 (**module**)。模块是 **Python** 组织代码的基本方式。

Python 的程序是由包 (**package**)、模块 (**module**) 和函数组成。包是由一系列模块组成的集合。模块是处理某一类问题的函数和类的集合。

包就是一个完成特定任务的工具箱，**Python** 提供了许多有用的工具包，如字符串处理、图形用户接口、WEB 接口、图形图像处理等。使用这些工具包，可以提高程序员的开发效率、减少编程的复杂度、达到代码重用的效果。这些自带的工具包和模块安装在 **Python** 的安装目录下的 **Lib** 子目录中。

例如，**Lib** 目录中的 **xml** 文件夹就是一个包，这个包用于完成 XML 的应用开发。**xml** 包中有几个子包：**dom**、**sax**、**etree** 和 **parser**。文件 **__init__.py**（注意是两个下滑线）是 **xml** 包的注册文件，如果没有该文件，**Python** 将不能识别 **xml** 包。在系统字典中定义了 **xml** 包。

包必须至少含有一个 **__init__.py** 文件，该文件的内容可以为空。**__init__.py** 用于表示当前文件夹是一个包。

1) 对于模块的理解

用简单的说法来说，每一个以扩展名为 **.py** 结尾的 **Python** 源代码都是一个模块。其他的文件可以通过导入一个模块读取这个模块的内容。导入从本质上来说，就是载入另一个文件，并能够读取那个文件的内容。一个模块的内容通过这样的属性能够被外部世界使用。

比如我们举一个简单的例子，首先建立一个 **python** 文件 **define.py**

```
-*-coding:utf-8-*-
myvar="这是一个测试"
```

很简单的代码就二行，我们定义了一个变量 **myvar**。现在我们通过另一个 **python** 文件来导入它。

```
-*-coding:utf-8-*-
import sys
import define
def Main():
    print define.myvar
    print dir(define)
    print define.__file__
if __name__=="__main__":
    Main()
```

```
import define #这一行语句的意义在于导入 define.py 这个模块
```

```
print define.myvar#这一行的意义在于打印 myvar 这个变量
```

```
print dir(define)#这一行的意义在于输出所有可以使用的变量
```

输出结果，不言而喻。但是在默认情况下，只是在每次会话的第一次运行。在第一次导入之后，其他的导入都不会再工作。

但是如果真的想要 `python` 在同一次会话中再次运行文件（不停止和重新启动会话），需要调用内置的 `reload` 函数。

2) 模块的显要特性：属性

导入和重载提供了一种自然的程序启动的选择，因为导入操作将会在最后一步执行文件。从更宏观的角度来看，模块扮演了一个工具库的角色。我们可以直接使用 `from define import myvar` 这条语句来实现。

```
-*-coding:utf-8-*-
import sys
from define import myvar
def Main():
    print myvar
if __name__=="__main__":
    Main()
```

3) 模块创建过程的例子

模板把一组相关的函数或代码组织到一个文件中。一个文件即是一个模板。模块由代码、函数或类组成。例如，创建一个名为 `myModule.py` 的文件，即定义了一个名为 `myModule` 的模块。在该模块中定义一个函数 `func()` 和一个类 `MyClass`。`MyClass` 类中定义一个方法 `myFunc()`。

```
-*-coding:utf-8-*-
import sys
def func():
    print "myModule.func()"
class MyClass:
    def myFunc(self):
        print "myModule.MyClass.myFunc()"
```

然后在 `myModule.py` 所在目录下创建一个名为 `call_myModule.py` 文件，在该文件中调用 `myModule` 模块的函数和类：

```
-*-coding:utf-8-*-
import sys
import random
import myModule
if __name__=="__main__":
    myModule.func()
    myClass=myModule.MyClass()
    myClass.myFunc()
```

当 `python` 寻入一个模块时，`python` 首先查找当前路径，然后查找 `lib` 目录、`site-packages` 目录 (`python/lib/site-packages`) 和环境变量 `PYTHONPATH` 设置的目录。

4) 模块的导入

在使用一个模块中的函数或类之前，首先要导入该模块。

```
import myModule
```

还可以使用 `from ...import..` 语句将模块导入。

```
from module_name import *
from module_name import function_name
```

实现例子

```
-*-coding:utf-8-*
import sys
def func():
    print "myModule.func()"
class MyClass:
    def myFunc(self):
        print "myModule.MyClass.myFunc()"
```

```
-*-coding:utf-8-*
import sys
import random
from myModule import func
if __name__=="__main__":
    func()
```

5) 模块的属性

模块中有许多内置的属性，用于完成特定的任务，如`__name__`、`__doc__`。每个模块都有一个名称。

```
-*-coding:utf-8-*
import sys
if __name__=="__main__":
    print __name__
    print __doc__
    print __file__
    print __package__
```

十五、 字符串与简单的正则表达式

本章将介绍 Python 中字符串和正则表达式的概念。字符串是程序开发中常用的数据类型，字符串的处理是实际应用中实际应用中经常面对的问题。Python 提供了功能强大的字符串模块，正则表达式专门用于匹配应用中的数据，能够简化字符串的处理程序，Python 提供了 `re` 模块用来匹配正则表达式。

如果我们有一个含有 12 个字符的字符串，我们通过内置的 `len` 函数验证其长度并通过索引操作得到其各个元

素。

在 Python 中，索引是按照从最前面的偏移量进行编码的，也就是从 0 开始，第一项索引为 0，第二项索引为 1，依次类推。在 python 中，我们能够反向索引，从最后一个开始。

```
>>> s='chuguangming'
>>> len(s)
12
>>> s[0]
'c'
>>> s[1]
'h'
>>> s[-1]
'g'
>>> s[-2]
'n'
>>>
```

除了简单的从位置进行索引，序列也支持一种所谓分片的操作，这是一种一步就能够提取整个分片的方法。例如：

```
>>> s
'chuguangming'
>>> s[1:3]
'hu'
>>> s[0:3]
'chu'
>>>

>>> s='chuguangming'
>>> s[1:]
'huguangming'
>>> s[0:3]
'chu'
>>> s[:3]
'chu'
>>> s[:1]
'c'
>>> s[:-1]
'chuguangmin'
>>> s[:]
'chuguangming'
>>>
```

```
>>> s='chu'
>>> s+'guangming'
'chuguangming'
>>> s*3
'chuchuchu'
>>>
```

1) 解决中文字符的问题

```
#-*- coding:utf-8 -*-
from __future__ import unicode_literals
print(type("test")) # output: <type 'unicode'>
```

```
spam = "测试字符串"  
print(spam[1:3]) # output:试字
```

2) 字符串的格式化

C 语言使用函数 `printf()`、`sprintf()` 格式化输出结果，Python 也提供了类似的功能。Python 将若干值插入到带有%标记的字符串，从而可以动态的输出字符串。字符串的格式化语法如下所示：

```
-*-coding:utf-8-*  
import sys  
if __name__=="__main__":  
    #格式化字符串  
    str1="version"  
    num=1.0  
    format="%s"%str1  
    print format  
    format="%s%d"%(str1,num)  
    print format
```

格式	描述
%%	百分号标记
%c	字符及其ASCII码
%s	字符串
%d	有符号整数(十进制)
%u	无符号整数(十进制)
%o	无符号整数(八进制)
%x	无符号整数(十六进制)
%X	无符号整数(十六进制大写字符)
%e	浮点数字(科学计数法)
%E	浮点数字(科学计数法，用E代替e)
%f	浮点数字(用小数点符号)
%g	浮点数字(根据值的大小采用%e或%f)
%G	浮点数字(类似于%g)
%p	指针(用十六进制打印值的内存地址)
%n	存储输出字符的数量放进参数列表的下一个变量中

```
-*-coding:utf-8-*-
```

```
import sys
if __name__=="__main__":
    #带精度的格式化字符串
    print "浮点型数字:%f"%1.25
    print "浮点型数字:%.1f"%1.25
    print "浮点型数字:%.2f"%1.254
```

```
-*-coding:utf-8-*
import sys
if __name__=="__main__":
    #带精度的格式化字符串
    print "浮点型数字:%f"%1.25
    print "浮点型数字:%.1f"%1.25
    print "浮点型数字:%.2f"%1.254

    #使用字典格式化字符串
    print "%(version)s: %(num).1f"%( {"version": "version", "num": 2})
    #字符串对齐
    word="version3.0"
    print word.center(20)
    print word.center(20, "*")
    #转义字符
    path="hello\tworld\n"
    print path
    print len(path)
    #直接输出转义
    path=r"hello\tworld\n"
    print path
```

3) 字符串的合并与分割

与 java 语言一样， Python 使用”+”号连接不同的字符串， Python 会根据加号两侧变量的类型，还决定连接操作或加法运算。

```
-*-coding:utf-8-*
import sys
if __name__=="__main__":
    #使用字符串截取子串
    word="world Hello"
    print word[0]
    #string[start:end:step]
    #从 0 开始截取三个
    print word[0:3]
    #从 0 开始
    print word[0::2]
    #使用 split() 获取子串
    sentence="Bob Said:1,2,3,4"
```

```
print "使用空格取子串: ", sentence.split()  
print "使用逗号取子串: ", sentence.split(',')  
print "使用两个逗号取子串: ", sentence.split(", ", 2)
```

```
-*- coding:utf-8-*  
import sys  
def Main():  
    print "string test example"  
    s='chuguangming'  
    #字符串替换  
    print s.find('chu')  
    print s.replace('chu','***')  
    #字符串分割  
    line='aaa,BBB,ccc,ddd'  
    result= line.split(',')  
    for s in result:  
        print s  
    #大小写转换  
    line=line.upper()  
    print line  
    #判断是不是字符串  
    print line.isalnum()  
    print line.isalpha()  
    #去除空格  
    line='aaa,bbb,ccc,ddd\n'  
    print line.rstrip()  
if __name__=="__main__":  
    Main()
```

正则表达式用于搜索、替换、解析字符串。正则表达式的功能强大，使用非常灵活。使用正则表达式需要遵守一定的语法规则，用来编写一些逻辑验证非常方便，例如电子邮件的验证。Python 提供了 `re` 模块实现正则表达式的验证。

```
 #-*-coding:utf-8-*  
import sys  
import re  
if __name__=="__main__":  
    #使用^与$的使用方法  
    s="HELLO WORLD"  
    #匹配 hello 开始的字符串，由于变量 s 中的 HELLO 是大写的所以匹配失败  
    print re.findall(r"^hello",s)  
    #re.I 表示时忽略大小写  
    print re.findall(r"^hello",s,re.I)  
    #$的意思是匹配尾部  
    print re.findall(r"WORD$",s)  
    #匹配每个英文单词\b 用于分割单词  
    print re.findall(r"\b\w+\b",s)
```

4) 常用字符操作演示

去空格及特殊符号

```
s.strip().lstrip().rstrip(',')
```

复制字符串

```
#strcpy(sStr1,sStr2)
sStr1 = 'strcpy'
sStr2 = sStr1
sStr1 = 'strcpy2'
print sStr2
```

连接字符串

```
#strcat(sStr1,sStr2)
sStr1 = 'strcat'
sStr2 = 'append'
sStr1 += sStr2
print sStr1
```

查找字符

```
#strchr(sStr1,sStr2)
# < 0 为未找到
sStr1 = 'strchr'
sStr2 = 's'
nPos = sStr1.index(sStr2)
print nPos
```

比较字符串

```
#strcmp(sStr1,sStr2)
sStr1 = 'strchr'
sStr2 = 'strch'
print cmp(sStr1,sStr2)
```

扫描字符串是否包含指定的字符

```
#strspn(sStr1,sStr2)
sStr1 = '12345678'
sStr2 = '456'
#sStr1 and chars both in sStr1 and sStr2
print len(sStr1 and sStr2)
```

字符串长度

```
#strlen(sStr1)
sStr1 = 'strlen'
print len(sStr1)
```

将字符串中的大小写转换

```
#strlwr(sStr1)
sStr1 = 'JCstrlwr'
sStr1 = sStr1.upper()
#sStr1 = sStr1.lower()
print sStr1
```

追加指定长度的字符串

```
#strncat(sStr1,sStr2,n)
sStr1 = '12345'
sStr2 = 'abcdef'
n = 3
sStr1 += sStr2[0:n]
print sStr1
```

字符串指定长度比较

```
#strcmp(sStr1,sStr2,n)
sStr1 = '12345'
sStr2 = '123bc'
n = 3
print cmp(sStr1[0:n],sStr2[0:n])
```

复制指定长度的字符

```
#strncpy(sStr1,sStr2,n)
sStr1 = ''
sStr2 = '12345'
n = 3
sStr1 = sStr2[0:n]
print sStr1
```

将字符串前 n 个字符替换为指定的字符

```
#strnset(sStr1,ch,n)
sStr1 = '12345'
ch = 'r'
n = 3
sStr1 = n * ch + sStr1[3:]
print sStr1
```

扫描字符串

```
#strupbrk(sStr1,sStr2)
sStr1 = 'cekjgdklab'
sStr2 = 'gka'
nPos = -1
for c in sStr1:
    if c in sStr2:
        nPos = sStr1.index(c)
        break
print nPos
```

翻转字符串

```
#strrev(sStr1)
sStr1 = 'abcdefg'
sStr1 = sStr1[::-1]
print sStr1
```

查找字符串

```
#strstr(sStr1,sStr2)
sStr1 = 'abcdefg'
```

```
sStr2 = 'cde'  
print sStr1.find(sStr2)
```

分割字符串

```
#strtok(sStr1,sStr2)  
sStr1 = 'ab,cde,fg,h,ijk'  
sStr2 = ','  
sStr1 = sStr1[sStr1.find(sStr2) + 1:]  
print sStr1  
#或者  
s = 'ab,cde,fg,h,ijk'  
print(s.split(','))
```

连接字符串

```
delimiter = ','  
mylist = ['Brazil', 'Russia', 'India', 'China']  
print delimiter.join(mylist)
```

PHP 中 addslashes 的实现

```
def addslashes(s):  
    d = {'\\': '\\\\', '\"': '\\"', '\\0': '\\\\0', '\\\\': '\\\\\\' }  
    return ''.join(d.get(c, c) for c in s)  
  
s = "John 'Johny' Doe (a.k.a. \"Super Joe\")\\\\\\0"  
print s  
print addslashes(s)
```

只显示字母与数字

```
def OnlyCharNum(s, oth=' '):  
    s2 = s.lower();  
    fomart = 'abcdefghijklmnopqrstuvwxyz0123456789'  
    for c in s2:  
        if not c in fomart:  
            s = s.replace(c, '');  
    return s;  
  
print(OnlyStr("a000 aa-b"))
```

十六、 文件 IO

1) 基本文件功能演示

```
#-*-coding:utf-8-*-
import sys
import re
if __name__=="__main__":
    #创建文件
    context='''
hello world
hello china '''
    f=file('hello.txt','w')
    f.write(context)
    f.close()
```

```
#-*-coding:utf-8-*-
import sys
import re
if __name__=="__main__":
    #读取文件
    f=open("hello.txt")
    while True:
        line=f.readline()
        if line:
            print line
        else:
            break
    f.close()
```

```
#-*-coding:utf-8-*-
import sys
import re
if __name__=="__main__":
    #使用 readlines()读取多个文件
    f=file("hello.txt")
    lines=f.readlines()
    print lines
    for line in lines:
        print line
```

```
#-*-coding:utf-8-*-
import sys
import re
if __name__=="__main__":
    #一次性读取
    f=open("hello.txt")
    context=f.read()
    print context
```

```
#-*-coding:utf-8-*-
import sys
```

```
import re
if __name__=="__main__":
    f=open("hello.txt")
    context=f.read(5)
    print context #读取前 5 个字节的数据
    print f.tell()#显示当前位置
    context=f.read(5)
    print context
    print f.tell()
    f.close()
```

2) 文件的写入

```
 #-*-coding:utf-8-*-
import sys
import re
if __name__=="__main__":
    #使用 writelines()写文件
    f=file("hello.txt","w+")
    li=["hello chu888\n","hello li\n"]
    f.writelines(li)
    f.close()
```

```
 #-*-coding:utf-8-*-
import sys
import re
if __name__=="__main__":
    #使用 writelines()写文件
    f=file("hello.txt","w+")
    li=["hello chu888 楚\n","hello li\n"]
    f.writelines(li)
    f.close()
    #追加文件内容
    f=file("hello.txt","a+")
    new_context="goodbye"
    f.write(new_context)
    f.close()
```

3) 文件的删除与复制

```
 #-*-coding:utf-8-*-
import sys
```

```
import re
import os
if __name__=="__main__":
    file("hello.txt","w")
    if os.path.exists("hello.txt"):
        os.remove("hello.txt")
```

`file` 类没有提供文件拷贝的功能，但是我们可以使用 `read()`、`write()` 方法模拟实现文件的拷贝，但是最好的方法是引入 `shutil` 模块

```
-*-coding:utf-8 -*-
import sys
import re
import os
import shutil
if __name__=="__main__":
    shutil.copyfile("a.txt","hello2.txt")
    shutil.move("a.txt","../")
    shutil.move("hello2.txt","aaa.txt")
```

4) 文件与目录的重命名

`os` 模块的函数 `rename()` 可以对文件或目录进行重命名

演示文件重命名的操作。如果当前目录存在名为 `hello.txt` 的文件，则重命名为 `hi.txt`；如果存在 `hi.txt` 的文件则重命名为 `hello.txt`

```
-*-coding:utf-8-
import sys
import os
if __name__=="__main__":
    li=os.listdir(".")#判断当前目录
    print li
    if "hello.txt" in li:
        os.rename("hello.txt","hi.txt")
    elif "hi.txt" in li:
        os.rename("hi.txt","hello.txt")
```

把后缀名为“`html`”的文件修改为“`htm`”后缀的文件

```
-*-coding:utf-8-
import sys
import os
```

```
if __name__=="__main__":
    files=os.listdir(".")
    for filename in files:
        #查找文件名中.所在的位置并把它给 pos
        pos=filename.find(".")
        #print pos
        #得到.后面的内容
        if filename[pos+1:]=="html":
            print filename
            newname=filename[:pos+1]+"htm"
            print newname
            os.rename(filename,newname)
```

```
-*-coding:utf-8-*
import sys
import os
if __name__=="__main__":
    files=os.listdir(".")
    for filename in files:
        li=os.path.splitext(filename)
        if li[1]==".html":
            newname=li[0]+".htm"
            os.rename(filename,newname)
```

5) 文件内容的查找和替换

从文件中查找字符串 hello，并统计 hello 出现的次数

```
-*-coding:utf-8-*
import sys
import os
import re
if __name__=="__main__":
    fl=file("aaa.txt","r")
    count=0
    for s in fl.readlines():
        li=re.findall("hello",s)
        if len(li)>0:
            count=count+li.count("hello")
    print "查找到"+str(count)+"个 hello"
    fl.close()
```

把 hello.txt 中的字符串 hello 全部替换为 hi，并把结果保存在文件 hello2.txt 中

```
-*-coding:utf-8-*
import sys
import os
```

```
import re
if __name__=="__main__":
    f1=file("aaa.txt","r")
    f2=file("bbb.txt","w")
    for s in f1.readlines():
        f2.write(s.replace("hello","hi"))
    f1.close()
    f2.close()
```

6) 文件的比较

Python 提供了模块 `difflib` 用于实现对序列、文件的比较。如果要比较两个文件，列出两个文件的异同，可以使用 `difflib` 模块的 `SequenceMatcher` 类实现。其中的方法 `get_opcodes()` 可以返回两个序列的比较结果。调用方法 `get_opcodes()` 之前，需要生成 1 个 `SequenceMatcher` 对象。

7) 控制台输入

`raw_input` 控制台输入

```
-*-coding:utf-8-*-
import os
myname = raw_input("please enter you name:")
print myname
```

`getpass.getpass` 方法

```
-*-coding:utf-8-*-
import os
import getpass
#简单
myname = raw_input("please enter you name:")
print myname
#输入密码安全
pwd = getpass.getpass('password: ')
print pwd
```

十七、 中文乱码

以下文章引用之：

<http://blog.csdn.net/kiki113/article/details/4062063>

<http://www.51testing.com/?uid-524463-action-viewspace-itemid-817888>

并适当编排

1) 编码普及

1. ASCII 码

我们知道，在计算机内部，所有的信息最终都表示为一个二进制的字符串。每一个二进制位（**bit**）有 **0** 和 **1** 两种状态，因此八个二进制位就可以组合出 **256** 种状态，这被称为一个字节（**byte**）。也就是说，一个字节一共可以用来表示 **256** 种不同的状态，每一个状态对应一个符号，就是 **256** 个符号，从 **00000000** 到 **11111111**。

上个世纪 **60** 年代，美国制定了一套字符编码，对英语字符与二进制位之间的关系，做了统一规定。这被称为 **ASCII 码**，一直沿用至今。

ASCII 码一共规定了 **128** 个字符的编码，比如空格“**SPACE**”是 **32**（二进制 **00100000**），大写的字母 **A** 是 **65**（二进制 **01000001**）。这 **128** 个符号（包括 **32** 个不能打印出来的控制符号），只占用了一个字节的后面 **7** 位，最前面的 **1** 位统一规定为 **0**。

2. 非 ASCII 编码

英语用 **128** 个符号编码就够了，但是用来表示其他语言，**128** 个符号是不够的。比如，在法语中，字母上方有注音符号，它就无法用 **ASCII 码** 表示。于是，一些欧洲国家就决定，利用字节中闲置的最高位编入新的符号。比如，法语中的 **é** 的编码为 **130**（二进制 **10000010**）。这样一来，这些欧洲国家使用的编码体系，可以表示最多 **256** 个符号。

但是，这里又出现了新的问题。不同的国家有不同的字母，因此，哪怕它们都使用 **256** 个符号的编码方式，代表的字母却不一样。比如，**130** 在法语编码中代表了 **é**，在希伯来语编码中却代表了字母 **Gimel** (ג)，在俄语编码中又会代表另一个符号。但是不管怎样，所有这些编码方式中，**0–127** 表示的符号是一样的，不一样的只是 **128–255** 的这一段。

至于亚洲国家的文字，使用的符号就更多了，汉字就多达 **10** 万左右。一个字节只能表示 **256** 种符号，肯定是不够的，就必须使用多个字节表达一个符号。比如，简体中文常见的编码方式是 **GB2312**，使用两个字节表示一个汉字，所以理论上最多可以表示 **256x256=65536** 个符号。

中文编码的问题需要专文讨论，这篇笔记不涉及。这里只指出，虽然都是用多个字节表示一个符号，但是 **GB** 类的汉字编码与后文的 **Unicode** 和 **UTF-8** 是毫无关系的。

3. Unicode

正如上一节所说，世界上存在着多种编码方式，同一个二进制数字可以被解释成不同的符号。因此，要想打开一个文本文件，就必须知道它的编码方式，否则用错误的编码方式解读，就会出现乱码。为什么电子邮件常常出现乱码？就是因为发信人和收信人使用的编码方式不一样。

可以想象，如果有一种编码，将世界上所有的符号都纳入其中。每一个符号都给予一个独一无二的编码，那么乱码问题就会消失。这就是 **Unicode**，就像它的名字都表示的，这是一种所有符号的编码。

Unicode 当然是一个很大的集合，现在的规模可以容纳 **100** 多万个符号。每个符号的编码都不一样，比如，**U+0639** 表示阿拉伯字母 **Ain**，**U+0041** 表示英语的大写字母 **A**，**U+4E25** 表示汉字“严”。具体的符号对应表，可以查询 **unicode.org**，或者专门的汉字对应表。

4. Unicode 的问题

需要注意的是，**Unicode** 只是一个符号集，它只规定了符号的二进制代码，却没有规定这个二进制代码应该如何存储。

比如，汉字“严”的 **unicode** 是十六进制数 **4E25**，转换成二进制数足足有 **15** 位（**100111000100101**），也就是

说这个符号的表示至少需要 2 个字节。表示其他更大的符号，可能需要 3 个字节或者 4 个字节，甚至更多。

这里就有两个严重的问题，第一个问题是，如何才能区别 **unicode** 和 **ascii**? 计算机怎么知道三个字节表示一个符号，而不是分别表示三个符号呢？第二个问题是，我们已经知道，英文字母只用一个字节表示就够了，如果 **unicode** 统一规定，每个符号用三个或四个字节表示，那么每个英文字母前都必然有二到三个字节是 0，这对于存储来说是极大的浪费，文本文件的大小会因此大出二三倍，这是无法接受的。

它们造成的结果是：1) 出现了 **unicode** 的多种存储方式，也就是说有许多种不同的二进制格式，可以用来表示 **unicode**。2) **unicode** 在很长一段时间内无法推广，直到互联网的出现。

5. UTF-8

互联网的普及，强烈要求出现一种统一的编码方式。**UTF-8** 就是在互联网上使用最广的一种 **unicode** 的实现方式。其他实现方式还包括 **UTF-16** 和 **UTF-32**，不过在互联网上基本不用。重复一遍，这里的关系是，**UTF-8** 是 **Unicode** 的实现方式之一。

UTF-8 最大的一个特点，就是它是一种变长的编码方式。它可以使用 1~4 个字节表示一个符号，根据不同的符号而变化字节长度。

UTF-8 的编码规则很简单，只有二条：

- 1) 对于单字节的符号，字节的第一位设为 0，后面 7 位为这个符号的 **unicode** 码。因此对于英文字母，**UTF-8** 编码和 **ASCII** 码是相同的。
- 2) 对于 n 字节的符号 ($n > 1$)，第一个字节的前 n 位都设为 1，第 $n+1$ 位设为 0，后面字节的前两位一律设为 10。剩下的没有提及的二进制位，全部为这个符号的 **unicode** 码。

下表总结了编码规则，字母 x 表示可用编码的位。

Unicode 符号范围 UTF-8 编码方式 (十六进制) (二进制)	
0000 0000-0000 007F	0XXXXXXX
0000 0080-0000 07FF	110XXXXX 10XXXXXX
0000 0800-0000 FFFF	1110XXXX 10XXXXXX 10XXXXXX
0001 0000-0010 FFFF	11110XXX 10XXXXXX 10XXXXXX 10XXXXXX

下面，还是以汉字“严”为例，演示如何实现 **UTF-8** 编码。

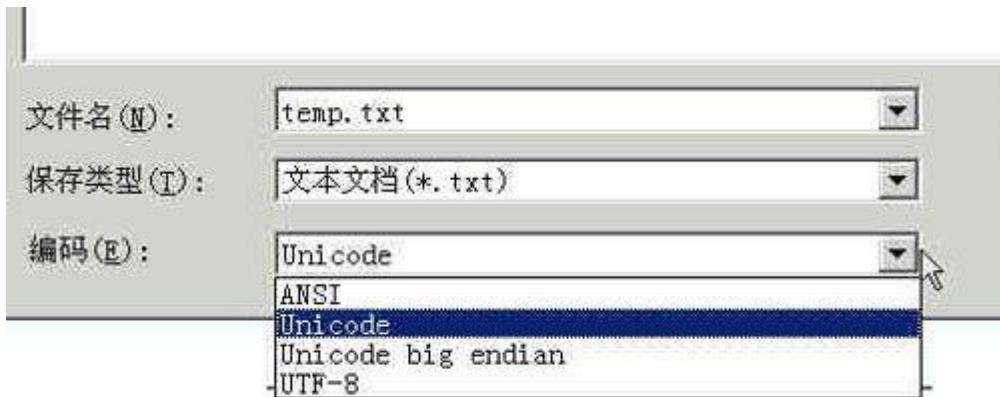
已知“严”的 **unicode** 是 4E25 (100111000100101)，根据上表，可以发现 4E25 处在第三行的范围内 (0000 0800-0000 FFFF)，因此“严”的 **UTF-8** 编码需要三个字节，即格式是“1110xxxx 10xxxxxx 10xxxxxx”。然后，从“严”的最后一个二进制位开始，依次从后向前填入格式中的 x，多出的位补 0。这样就得到了，“严”的 **UTF-8** 编码是“11100100 10111000 10100101”，转换成十六进制就是 E4B8A5。

6. Unicode 与 UTF-8 之间的转换

通过上一节的例子，可以看到“严”的 **unicode** 码是 4E25，**UTF-8** 编码是 E4B8A5，两者是不一样的。它们之间的转换可以通过程序实现。

在 **Windows** 平台下，有一个最简单的转化方法，就是使用内置的记事本小程序 **Notepad.exe**。打开文件后，

点击“文件”菜单中的“另存为”命令，会跳出一个对话框，在最底部有一个“编码”的下拉条。



里面有四个选项：ANSI, Unicode, Unicode big endian 和 UTF-8。

1) ANSI 是默认的编码方式。对于英文文件是 ASCII 编码，对于简体中文文件是 GB2312 编码（只针对 Windows 简体中文版，如果是繁体中文版会采用 Big5 码）。

2) Unicode 编码指的是 UCS-2 编码方式，即直接用两个字节存入字符的 Unicode 码。这个选项用的 little endian 格式。

3) Unicode big endian 编码与上一个选项相对应。我在下一节会解释 little endian 和 big endian 的涵义。

4) UTF-8 编码，也就是上一节谈到的编码方法。

选择完“编码方式”后，点击“保存”按钮，文件的编码方式就立刻转换好了。

7. Little endian 和 Big endian

上一节已经提到，Unicode 码可以采用 UCS-2 格式直接存储。以汉字“严”为例，Unicode 码是 4E25，需要用两个字节存储，一个字节是 4E，另一个字节是 25。存储的时候，4E 在前，25 在后，就是 Big endian 方式；25 在前，4E 在后，就是 Little endian 方式。

这两个古怪的名称来自英国作家斯威夫特的《格列佛游记》。在该书中，小人国里爆发了内战，战争起因是人们争论，吃鸡蛋时究竟是从大头(Big-Endian)敲开还是从小头(Little-Endian)敲开。为了这件事情，前后爆发了六次战争，一个皇帝送了命，另一个皇帝丢了王位。

因此，第一个字节在前，就是“大头方式”(Big endian)，第二个字节在前就是“小头方式”(Little endian)。那么很自然的，就会出现一个问题：计算机怎么知道某一个文件到底采用哪一种方式编码？

Unicode 规范中定义，每一个文件的最前面分别加入一个表示编码顺序的字符，这个字符的名字叫做“零宽度非换行空格”(ZERO WIDTH NO-BREAK SPACE)，用 FFFF 表示。这正好是两个字节，而且 FF 比 FE 大 1。

如果一个文本文件的头两个字节是 FE FF，就表示该文件采用大头方式；如果头两个字节是 FF FE，就表示该文件采用小头方式。

8. 实例

下面，举一个实例。

打开“记事本”程序 Notepad.exe，新建一个文本文件，内容就是一个“严”字，依次采用 ANSI, Unicode, Unicode big endian 和 UTF-8 编码方式保存。

然后，用文本编辑软件 UltraEdit 中的“十六进制功能”，观察该文件的内部编码方式。

1) ANSI：文件的编码就是两个字节“D1 CF”，这正是“严”的 GB2312 编码，这也暗示 GB2312 是采用大头方式存储的。

2) Unicode：编码是四个字节“FF FE 25 4E”，其中“FF FE”表明是小头方式存储，真正的编码是 4E25。

3) Unicode big endian：编码是四个字节“FE FF 4E 25”，其中“FE FF”表明是大头方式存储。

4) UTF-8：编码是六个字节“EF BB BF E4 B8 A5”，前三个字节“EF BB BF”表示这是 UTF-8 编码，后三个“E4B8A5”

2) 中文乱码细究

在本文中，以'哈'来解释作示例解释所有的问题，“哈”的各种编码如下：

1. UNICODE (UTF8-16), C854;
2. UTF-8, E59388;
3. GBK, B9FE。

1. python 中的 str 和 unicode

一直以来，python 中的中文编码就是一个极为头大的问题，经常抛出编码转换的异常，python 中的 `str` 和 `unicode` 到底是一个什么东西呢？

在 python 中提到 `unicode`，一般指的是 `unicode` 对象，例如'哈哈'的 `unicode` 对象为 `u'/u54c8/u54c8'` 而 `str`，是一个字节数组，这个字节数组表示的是对 `unicode` 对象编码(可以是 utf-8、gbk、cp936、GB2312)后的存储的格式。这里它仅仅是一个字节流，没有其它的含义，如果你想使这个字节流显示的内容有意义，就必须用正确的编码格式，解码显示。例如：

```
>>> su = u'哈哈'
>>> s_utf8 = su.encode('utf-8')
>>> s_utf8
'\xe5\x93\x88\xe5\x93\x88'
>>> print s_utf8
鍋堢握手
>>> s_gbk = su.encode('gbk')
>>> s_gbk
'\xb9\xfe\xb9\xfe'
>>> print s_gbk
哈哈
>>> print su
哈哈
>>> _
```

对于 `unicode` 对象哈哈进行编码，编码成一个 `utf-8` 编码的 `str`—`s_utf8`,`s_utf8` 就是一个字节数组，存放的就是'`/xe5/x93/x88/xe5/x93/x88`'，但是这仅仅是一个字节数组，如果你想将它通过 `print` 语句输出成哈哈，那你就失望了，为什么呢？

因为 `print` 语句它的实现是将要输出的内容传送了操作系统，操作系统会根据系统的编码对输入的字节流进行编码，这就解释了为什么 `utf-8` 格式的字符串“哈哈”，输出的是“鍋堢握手”，因为'`/xe5/x93/x88/xe5/x93/x88`'用 `GB2312` 去解释，其显示的出来就是“鍋堢握手”。这里再强调一下，`str` 记录的是字节数组，只是某种编码的存储格式，至于输出到文件或是打印出来是什么格式，完全取决于其解码的编码将它解码成什么样子。

这里再对 `print` 进行一点补充说明：当将一个 `unicode` 对象传给 `print` 时，在内部会将该 `unicode` 对象进行一次转换，转换成本地的默认编码（这仅是个人猜测）

2. str 和 unicode 对象的转换

`str` 和 `unicode` 对象的转换，通过 `encode` 和 `decode` 实现，具体使用如下：

```
>>> s = '哈哈'  
>>> print s.decode('gbk').encode('utf-8')  
鍋堄撢
```

将 GBK '哈哈' 转换成 unicode，然后再转换成 UTF8

3) Setdefaultencoding

```
Type "help", "copyright", "credits" or "license" for more information.  
>>> s = '哈哈'  
>>> s.encode('utf-8')  
Traceback (most recent call last):  
  File "<stdin>", line 1, in <module>  
UnicodeDecodeError: 'ascii' codec can't decode byte 0xb9 in position 0: ordinal  
not in range<128>  
>>> import sys  
>>> reload(sys)  
<module 'sys' (built-in)>  
>>> sys.setdefaultencoding('gbk')  
>>> s.encode('utf-8')  
'\xe5\x93\x88\xe5\x93\x88'  
>>>
```

如上图的演示代码所示：当把 s(gbk 字符串)直接编码成 utf-8 的时候，将抛出异常，但是通过调用如下代码：

```
import sys  
reload(sys)  
sys.setdefaultencoding('gbk')
```

后就可以转换成功，为什么呢？在 python 中 str 和 unicode 在编码和解码过程中，如果将一个 str 直接编码成另一种编码，会先把 str 解码成 unicode，采用的编码为默认编码，一般默认编码是 ascii，所以在上面示例代码中第一次转换的时候会出错，当设定当前默认编码为 'gbk' 后，就不会出错了。至于 reload(sys) 是因为 Python2.5 初始化后会删除 sys.setdefaultencoding 这个方法，我们需要重新载入。

3. 操作不同文件的编码格式的文件

建立一个文件 test.txt，文件格式用 ANSI，内容为：abc 中文，用 python 来读取。

```
# coding=gbk  
print open("Test.txt").read()
```

结果： abc 中文

把文件格式改成 UTF-8：

结果： abc 涓 构

显然，这里需要解码：

```
# coding=gbk  
import codecs  
print open("Test.txt").read().decode("utf-8")
```

结果： abc 中文

上面的 test.txt 我是用 Editplus 来编辑的，但当我用 Windows 自带的记事本编辑并存成 UTF-8 格式时，运行时报错：

```
Traceback (most recent call last):
File "ChineseTest.py", line 3, in
print open("Test.txt").read().decode("utf-8")
UnicodeEncodeError: 'gbk' codec can't encode character u'/ufeef' in position 0: illegal
multibyte sequence
```

原来，某些软件，如 **notepad**，在保存一个以 **UTF-8** 编码的文件时，会在文件开始的地方插入三个不可见的字符（**0xEF 0xBB 0xBF**，即 **BOM**）。

因此我们在读取时需要自己去掉这些字符，**python** 中的 **codecs module** 定义了这个常量：

```
# coding=gbk
import codecs
data = open("Test.txt").read()
if data[:3] == codecs.BOM_UTF8:
    data = data[3:]
print data.decode("utf-8")
```

结果：abc 中文

4. 文件的编码格式和编码声明的作用

源文件的编码格式对字符串的声明有什么作用呢？这个问题困扰一直困扰了我好久，现在终于有点眉目了，文件的编码格式决定了在该源文件中声明的字符串的编码格式，例如：

```
str = '哈哈'
print repr(str)
```

- a.如果文件格式为 **utf-8**，则 **str** 的值为：'**/xe5/x93/x88/xe5/x93/x88**'
- b.如果文件格式为 **gbk**，则 **str** 的值为：'**/xb9/xfe/xb9/xfe**'（哈哈的 **gbk** 编码）

在第一节已经说过，**python** 中的字符串，只是一个字节数组，所以当把 a 情况的 **str** 输出到 **gbk** 编码的控制台时，就将显示为乱码：鐸堦搘；而当把 b 情况下的 **str** 输出 **utf-8** 编码的控制台时，也将显示乱码的问题，是什么也没有，也许'**/xb9/xfe/xb9/xfe**'用 **utf-8** 解码显示，就是空白吧。>_<

说完文件格式，现在来谈谈编码声明的作用吧，每个文件在最上面的地方，都会用# **coding=gbk** 类似的语句声明一下编码，但是这个声明到底有什么用呢？到目前为止，我觉得它的作用也就是三个：

- 声明源文件中将出现非 **ascii** 编码，通常也就是中文；
- 在高级的 **IDE** 中，**IDE** 会将你的文件格式保存成你指定编码格式。
- 决定源码中类似于 **u'哈'** 这类声明的将‘哈’解码成 **unicode** 所用的编码格式，也是一个比较容易让人迷惑的地方，看示例：

```
#coding:gbk
ss = u'哈哈'
print repr(ss)
print 'ss:%s' % ss
```

将这个些代码保存成一个 **utf-8** 文本，运行，你认为会输出什么呢？大家第一感觉肯定输出的肯定是：

```
u'/u54c8/u54c8'
ss:哈哈
```

但实际上输出是：

```
u'/u935d/u581d/u6431'
ss:鐸堦搘
```

为什么会这样，这时候，就是编码声明在作怪了，在运行 **ss = u'哈哈'** 的时候，整个过程可以分为以下几步：

- 1) 获取‘哈’的编码：由文件编码格式确定，为'**/xe5/x93/x88/xe5/x93/x88**'（哈哈的 **utf-8** 编码形式）

2) 转成 `unicode` 编码的时候，在这个转换的过程中，对于'xe5/x93/x88/xe5/x93/x88'的解码，不是用 `utf-8` 解码，而是用声明编码处指定的编码 `GBK`，将'xe5/x93/x88/xe5/x93/x88'按 `GBK` 解码，得到就是“鍋堢搣”，这三个字的 `unicode` 编码就是 `u'/u935d/u581d/u6431'`，至此可以解释为什么 `print repr(ss)` 输出的是 `u'/u935d/u581d/u6431'` 了。

好了，这里有点绕，我们来分析下一个示例：

```
#-*- coding:utf-8 -*-
ss = u'哈哈'
print repr(ss)
print 'ss:%s' % ss
```

将这个示例这次保存成 `GBK` 编码形式，运行结果，竟然是：

```
UnicodeDecodeError: 'utf8' codec can't decode byte 0xb9 in position 0: unexpected code
byte
```

4) 新手编码必读 (gonewiththewind 大虾作品)

为什么会报错

```
"UnicodeEncodeError: 'ascii' codec can't encode characters in position 0-1: ordinal not in
range(128)"
```

本文就来研究一下这个问题。字符串在 `Python` 内部的表示是 `unicode` 编码，因此，在做编码转换时，通常需要以 `unicode` 作为中间编码，即先将其他编码的字符串解码(`decode`)成 `unicode`，再从 `unicode` 编码(`encode`)成另一种编码。

`decode` 的作用是将其他编码的字符串转换成 `unicode` 编码，如 `str1.decode('gb2312')`，表示将 `gb2312` 编码的字符串 `str1` 转换成 `unicode` 编码。

`encode` 的作用是将 `unicode` 编码转换成其他编码的字符串，如 `str2.encode('gb2312')`，表示将 `unicode` 编码的字符串 `str2` 转换成 `gb2312` 编码。

因此，转码的时候一定要先搞明白，字符串 `str` 是什么编码，然后 `decode` 成 `unicode`，然后再 `encode` 成其他编码，代码中字符串的默认编码与代码文件本身的编码一致。

如：`s='中文'`

如果是在 `utf8` 的文件中，该字符串就是 `utf8` 编码，如果是在 `gb2312` 的文件中，则其编码为 `gb2312`。这种情况下，要进行编码转换，都需要先用 `decode` 方法将其转换成 `unicode` 编码，再使用 `encode` 方法将其转换成其他编码。通常，在没有指定特定的编码方式时，都是使用的系统默认编码创建的代码文件。

如果字符串是这样定义：`s=u'中文'`，则该字符串的编码就被指定为 `unicode` 了，即 `python` 的内部编码，而与代码文件本身的编码无关。因此，对于这种情况做编码转换，只需要直接使用 `encode` 方法将其转换成指定编码即可。

如果一个字符串已经是 `unicode` 了，再进行解码则将出错，因此通常要对其编码方式是否为 `unicode` 进行判断：

```
isinstance(s, unicode)#用来判断是否为 unicode, 用非 unicode 编码形式的 str 来 encode 会报错。
```

如何获得系统的默认编码？

```
#!/usr/bin/env python
#coding=utf-8
import sys
print sys.getdefaultencoding()
```

该段程序在英文 `WindowsXP` 上输出为：`ascii`

在某些 `IDE` 中，字符串的输出总是出现乱码，甚至错误，其实是由于 `IDE` 的结果输出控制台自身不能显示字符串的编码，而不是程序本身的问题。

如在 `UliPad` 中运行如下代码：

```
s=u"中文"
```

```
print s
```

会提示: `UnicodeEncodeError: 'ascii' codec can't encode characters in position 0-1: ordinal not in range(128)`。

这是因为 ULiPad 在英文 WindowsXP 上的控制台信息输出窗口是按照 `ascii` 编码输出的（英文系统的默认编码是 `ascii`），而上面代码中的字符串是 `Unicode` 编码的，所以输出时产生了错误。

将最后一句改为: `print s.encode('gb2312')` 则能正确输出“中文”两个字。若最后一句改为:

```
print s.encode('utf8')
```

则输出: `\xe4\xb8\xad\xe6\x96\x87`, 这是控制台信息输出窗口按照 `ascii` 编码输出 `utf8` 编码的字符串的结果。

`unicode(str, 'gb2312')` 与 `str.decode('gb2312')` 是一样的，都是将 `gb2312` 编码的 `str` 转为 `unicode` 编码。使用 `str.__class__` 可以查看 `str` 的编码形式

十八、 Web.Py 框架概述

1) Python 下 web 开发框架的选择

以下内容为网络摘录，来源：飞龙博客

1. Django



Python 框架虽然说是百花齐放，但仍然有那么一家是最大的，它就是 `Django`。要说 `Django` 是 Python 框架里最好的，有人同意也有人坚决反对，但说 `Django` 的文档最完善、市场占有率最高、招聘职位最多估计大家都没什么意见。`Django` 为人所称道的地方主要有：

- 完美的文档, `Django` 的成功, 我觉得很大一部分原因要归功于 `Django` 近乎完美的官方文档(包括 `Django book`)。
- 全套的解决方案, `Django` 象 `Rails` 一样, 提供全套的解决方案 (`full-stack framework + batteries included`), 基本要什么有什么 (比如: `cache`、`session`、`feed`、`orm`、`geo`、`auth`), 而且全部 `Django` 自己造, 开发网站应手的工具 `Django` 基本都给你做好了, 因此开发效率是不用说的, 出了问题也算好找, 不在你的代码里就在 `Django` 的源码里。
- 强大的 URL 路由配置, `Django` 让你可以设计出非常优雅的 URL, 在 `Django` 里你基本可以跟丑陋的 GET 参数说拜拜。
- 自助管理后台, `admin interface` 是 `Django` 里比较吸引眼球的一项 `contrib`, 让你几乎不用写一行代码就拥有一个完整的后台管理界面。

而 `Django` 的缺点主要源自 `Django` 坚持自己造所有的轮子, 整个系统相对封闭, `Django` 最为人诟病的地方有:

- 系统紧耦合, 如果你觉得 `Django` 内置的某项功能不是很好, 想用喜欢的第三方库来代替是很难的, 比如下面将要说的 `ORM`、`Template`。要在 `Django` 里用 `SQLAlchemy` 或 `Mako` 几乎是不可能, 即使打了一些补丁用上了也会让你觉得非常非常别扭。

- Django 自带的 ORM 远不如 SQLAlchemy 强大，除了在 Django 这一亩三分地，SQLAlchemy 是 Python 世界里事实上的 ORM 标准，其它框架都支持 SQLAlchemy 了，唯独 Django 仍然坚持自己的那一套。Django 的开发人员对 SQLAlchemy 的支持也是有过讨论和尝试的，不过最终还是放弃了，估计是代价太高且跟 Django 其它的模块很难合到一块。
- Template 功能比较弱，不能插入 Python 代码，要写复杂一点的逻辑需要另外用 Python 实现 Tag 或 Filter。关于模板这一点，一直以来争论比较多，最近有两篇关于 Python 模板的比较有意思的文章可供参考：
 1. <http://pydanny.blogspot.com/2010/12/stupid-template-languages.html> (需翻墙)
 2. <http://techspot.zzeek.org/2010/12/04/in-response-to-stupid-template-languages/>
- URL 配置虽然强大，但全部要手写，这一点跟 Rails 的 Convention over configuration 的理念完全相左，高手和初识 Django 的人配出来的 URL 会有很大差异。
- 让人纠结的 auth 模块，Django 的 auth 跟其它模块结合紧密，功能也挺强的，就是做的有点过了，用户的数据库 schema 都给你定好了，这样问题就来了，比如很多网站要求 email 地址唯一，可 schema 里这个字段的值不是唯一的，纠结是必须的了。
- Python 文件做配置文件，而不是更常见的 ini、xml 或 yaml 等形式。这本身不是什么问题，可是因为理论上来说 settings 的值是能够动态的改变的（虽然大家不会这么干），但这不是最佳实践的体现。

总的来说，Django 大包大揽，用它来快速开发一些 Web 运用是很不错的。如果你顺着 Django 的设计哲学来，你会觉得 Django 很好用，越用越爽；相反，你如果不能融入或接受 Django 的设计哲学，你用 Django 一定会很痛苦，趁早放弃的好。所以说在有些人眼里 Django 无异于仙丹，但对有一些人来说它又是毒药且剧毒。

Django 案例有 disqus.com、bitbucket.org、[海报网](https://haibao.net) 等。

2. Pylons & TurboGears & repoze.bfg



除了 Django 另一个大头就是 Pylons 了，因为 TurboGears2.x 是基于 Pylons 来做的，而 repoze.bfg 也已经并入 Pylons project 里这个大的项目里，后面不再单独讨论 TurboGears 和 repoze.bfg 了。

Pylons 和 Django 的设计理念完全不同，Pylons 本身只有两千行左右的 Python 代码，不过它还附带有一些几乎就是 Pylons 御用的第三方模块。Pylons 只提供一个架子和可选方案，你可以根据自己的喜好自由的选择 Template、ORM、form、auth 等组件，系统高度可定制。我们常说 Python 是一个胶水语言(glue language)，那么我们完全可以说 Pylons 就是一个用胶水语言设计的胶水框架：)

选择 Pylons 多是选择了它的自由，选择了自由的同时也预示着你选择了噩梦：

- 学习噩梦，Pylons 依赖于许多第三方库，它们并不是 Pylons 造，你学 Pylons 的同时还得学这些库怎么使用，关键有些时候你都不知道你要学什么。Pylons 的学习曲线相对比 Django 要高的多，而之前 Pylons 的官方文档也一直是人批评的对象，好在后来出了 [The Definitive Guide to Pylons](#) 这本书，这一局面有所改观。因为这个原因，Pylons 一度被誉为只适合高手使用的 Python 框架。

- 调试噩梦，因为牵涉到的模块多，一旦有错误发生就比较难定位问题在哪里。可能是你写的程序的错、也可能 Pylons 出错了、再或是 SQLAlchemy 出错了、搞不好是 formencode 有 bug，反正很凌乱了。这个只有用的很熟了才能解决这个问题。
- 升级噩梦，安装 Pylons 大大小小共要安装近 20 个 Python 模块，各有各自的版本号，要升级 Pylons 的版本，哪个模块出了不兼容的问题都有可能，升级基本上很难很难。至今 reddit 的 Pylons 还停留在古董的 0.9.6 上，SQLAlchemy 也还是 0.5.3 的版本，应该跟这条有关系。所以大家玩 Pylons 一定要结合 virtualenv 来玩，给自己留条后路，不然会死得很惨。

Pylons 和 repoze.bfg 的融合可能会催生下一个能挑战 Django 地位的框架。

Pylons 的案例有 reddit.com、dropbox.com、quora.com 等。

3. Tornado & web.py



Tornado 即是一个 web server(对此本文不作详述)，同时又是一个类 web.py 的 micro-framework，作为框架 Tornado 的思想主要来源于 web.py，大家在 web.py 的网站首页也可以看到 Tornado 的大佬 [Bret Taylor](#) 的这么一段话（他这里说的 FriendFeed 用的框架跟 Tornado 可以看作是一个东西）：

“[web.py inspired the] web framework we use at FriendFeed [and] the webapp framework that ships with App Engine...”

因为有这层关系，后面不再单独讨论 Tornado。

web.py 的设计理念力求精简（Keep it simple and powerful），总共就没多少行代码，也不像 Pylons 那样依赖大量的第三方模块，而是只提供的一个框架所必须的一些东西，如：URL 路由、Template、数据库访问，其它的就交给用户自己去做好了。

一个框架精简的好处在于你可以聚焦在业务逻辑上，而不用太多的去关心框架本身或受框架的干扰，同时缺点也很明显，许多事情你得自己操刀上。

我个人比较偏好这种精简的框架，因为你很容易通过阅读源码弄明白整个框架的工作机制，如果框架那一块不是很合意的话，我完全可以 Monkey patch 一下按自己的要求来。

早期的 reddit 是用 web.py 写的，Tornado 的案例有 friendfeed.com、bit.ly、quora.com 和我的开源站点 poweredsites.org 等。

4. Bottle & Flask



Bottle 和 Flask 作为新生一代 Python 框架的代表，挺有意思的是都采用了 `decorator` 的方式配置 URL 路由，如：

```
from bottle import route, run

@route('/:name')
def index(name='World'):
    return '<Hello %s!>' % name

run(host='localhost', port=8080)
```

Bottle、Flask 跟 `web.py` 一样，都非常精简，Bottle 甚至所有的代码都在那一个两千来行的 `.py` 文件里。另外 Flask 和 Pylons 一样，可以跟 `Jinja2`、`SQLAlchemy` 之类结合的很好。

不过目前不管是 Bottle 还是 Flask 成功案例都还很少。

5. Quixote

之所以要特别说一下 `Quixote`，是因为国内的最大的用 Python 开发的网站“[豆瓣网](#)”是用 `Quixote` 开发的。我只简单翻了一下源代码，没有做过研究，不发表评论，有经验的来补充下。我只是在想，如果豆瓣网交到现在来开发，应该会有更多的选择。

6. 最后关于框架选择的误区

在框架的选择问题上，许多人很容易就陷入了下面两个误区中而不自知：

1. 哪个框架最好 — 世上没有最好的框架，只有最适合你自己、最适合你的团队的框架。编程语言选择也是一个道理，你的团队 Python 最熟就用 Python 好了，如果最熟悉的是 Ruby 那就用 Ruby 好了，编程语言、框架都只是工具，能多、快、好、省的干完活就是好东西，管 TMD 是日本鬼子还是美帝造呢！
2. 过分关注性能 — 其实大部分人是没必要太关心框架的性能的，因为你开发的网站根本就是个小站，能上 1 万的 IP 的网站已经不多了，上 10 万的更是很少很少。在没有一定的访问量前谈性能其实是没有多大意义的，因为你的 CPU 和内存一直就闲着呢。而且语言和框架一般也不会是性能瓶颈，性能问题最常出现在数据库访问和文件读写上。`PHP` 的 Zend Framework 是出了名的慢，但是 Zend Framework 一样有大站，如： [digg.com](#)；常被人说有性能问题的 Ruby 和 Rails，不是照样可以开发出 `twitter` 吗？再者现在的硬件、带宽成本其实是很低的，特别有了云计算平台后，人力成本才是最贵的，没有上万的 IP 根本就不用太在意性能问题，流量上去了花点钱买点服务器空间好了，简单快速的解决性能问题。

注：前面有网友质疑我“`Quora` 是用 `Pylons` 开发的”这样的说法不客观，特说明一下，这里所说的某个网站 A 是用 B 开发的，只是指 A 主要或部分是由 B 开发的，大家就不要再去纠结 A 还用 C 了。

2) 解决 SAE 第三方包的安装方案

1、安装 `virtualenv`

```
Microsoft Windows [版本 6.1.7601]
版权所有 (c) 2009 Microsoft Corporation。保留所有权利。

C:\Users\TenYear>cd ..

C:\Users>cd ..

C:\>easy_install virtualenv
Searching for virtualenv
Reading http://pypi.python.org/simple/virtualenv/
Reading http://www.virtualenv.org
Reading http://virtualenv.openplans.org
```

2、安装 pip

```
C:\>easy_install pip
Searching for pip
Reading http://pypi.python.org/simple/pip/
Reading http://www.pip-installer.org
Reading http://pip.openplans.org
```

3、创建一个 Python 的虚拟环境，不依赖于本机的 Python 原有环境

```
C:\PythonProject>virtualenv --no-site-packages ENV
New python executable in ENV\Scripts\python.exe
Installing setuptools.....done.
Installing pip.....done.
```

4、这时会在当前目录生成一个 ENV 目录，进入这个目录中

```
C:\PythonProject\ENV\Scripts 的目录

[.]          [...]          activate
activate.ps1  activate_this.py  deactivate.bat
easy_install-2.7.exe  easy_install-2.7.exe.manifest  easy_install-script.py
easy_install.exe.manifest  pip-2.7-script.py  pip-2.7.exe
pip-script.py  pip.exe  pip.exe.manifest
pythonw.exe

19 个文件      99,326 字节
2 个目录 250,047,995,904 可用字节

C:\PythonProject\ENV\Scripts>activate.bat
(ENV) C:\PythonProject\ENV\Scripts>
```

5、下载所需要打包的第三方包，并且冻结

```
C:\PythonProject\ENV\Scripts>activate.bat
(ENV) C:\PythonProject\ENV\Scripts>pip install BeautifulSoup
Downloading/unpacking BeautifulSoup
  Downloading BeautifulSoup-3.2.1.tar.gz
  Running setup.py egg_info for package BeautifulSoup

Installing collected packages: BeautifulSoup
  Running setup.py install for BeautifulSoup

Successfully installed BeautifulSoup
Cleaning up...

(ENV) C:\PythonProject\ENV\Scripts>pip freeze > requirements.txt
(ENV) C:\PythonProject\ENV\Scripts>
```

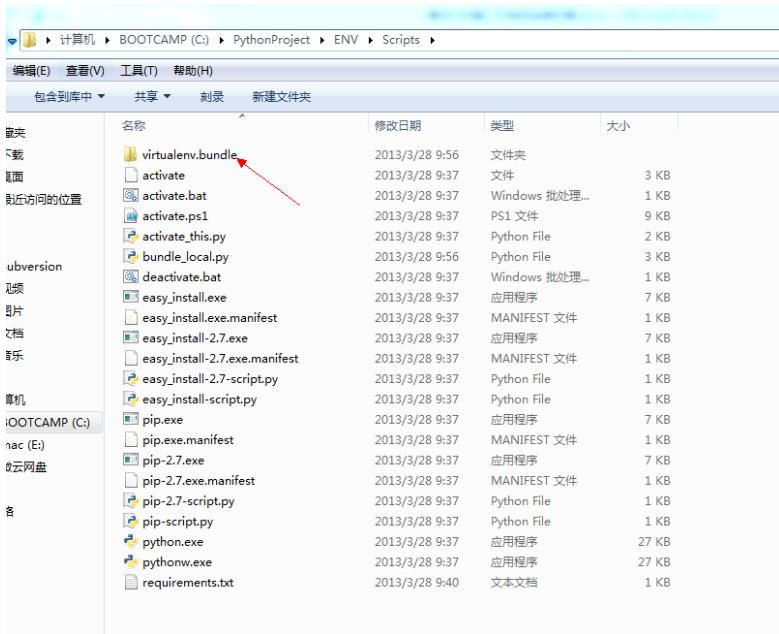
6、需要去 Github 下载服务器打包工具。

https://github.com/SAEPython/saepythondevguide/tree/master/dev_server

sae	a month ago	Add support for tornado to dev_server [chanfung032]
README	11 months ago	Add docs for sae.kvdb [chanfung032]
bundle_local.py	4 months ago	handle .egg package [chanfung032]
dev_server.py	a month ago	Add support for tornado to dev_server [chanfung032]
saecloud	16 days ago	Add -r option for `saecloud install` [chanfung032]
saecloud.bat	4 months ago	Fix windows compatibility [chanfung032]
setup.py	15 days ago	Update meta info [chanfung032]

7、打包

```
python bundle_local.py -r requirements.txt
```



将 virtualenv.bundle 目录到应用的目录下。修改 index.wsgi 文件，在导入其它模块之前，将 virtualenv.bundle 目录或者 virtualenv.bundle.zip 添加到 module 的搜索路径中，示例代码如下：

```
import os
import sys
app_root = os.path.dirname(__file__)
# 两者取其一
sys.path.insert(0,os.path.join(app_root,'virtualenv.bundle'))
```

到此，所有的依赖包已经导出并加入到应用的目录里了。

3) web.py 安装

首先下载 Web.Py 的安装包：

```
http://webpy.org/static/web.py-0.37.tar.gz
```

或者从 GITHUB 下载最新的安装包也可以：

```
https://github.com/webpy/webpy/tarball/master
```

使用命令行安装

```
python setup.py install
```

如果是 Linux 下面的话，需要使用 Sudo

```
sudo python setup.py install
```

4) web.py 下常用框架简介

- **jinja2**, **python** 中一个很流行的模板框架, 用了它, 写模板就是如此的享受。
- **sqlalchemy**, **python** 最强大的 **orm**, 没有之一, 掌握了它, 数据库? **so easy!**
- **formalchemy**, 配套 **sqlalchemy** 的表单框架, 可以根据 **sqlalchemy** 中数据表的定义生成 **html** 表单
- **kendo ui**, 这是一个 **jquery** 框架, 提供了很多插件, 里面最让我欣赏的, 就是 **mvvm** 的开发模式和 **datasource** 的 **ajax**, 让本来已经很少的 **jquery** 代码变得更少了。

5) 最简单的 HelloWorld!

```
#-*- coding:utf-8 -*-
import web

#定义 url, 将地址映射到对应的类
urls = (
    "/", "index",
)

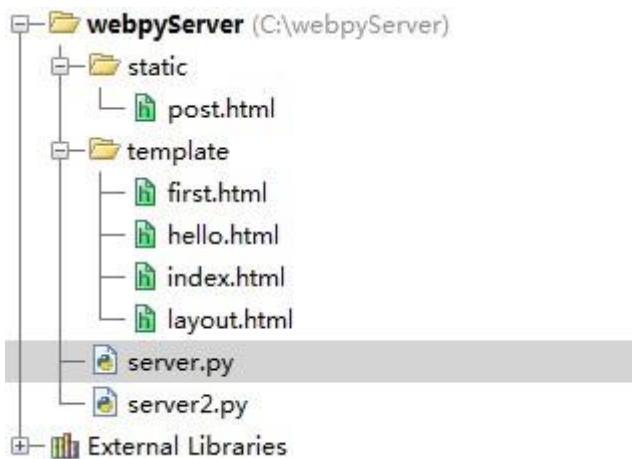
app = web.application(urls, globals())

#定义 index 类
class index:
    #get 请求
    def GET(self):
        return "Hello World"

if __name__ == "__main__":
    app.run()
```

6) Web.py 基本使用模板使用

项目结构



server.py

```
#-*-coding:utf-8-*-
import web
urls = (
    '/', 'index'
)
render = web.template.render('template/')
class index:
    def GET(self):
        #return render.hello('楚广明')
        return render.first('world', 'name')
if __name__ == "__main__":
    app = web.application(urls, globals())
    app.run()
```

first.html

```
$def with (name1,name2)
<html><head>
    <title>my si

    </title>
</head><body>
Hello $name1 , hello $name2!
<table>
    $for c in ["a", "b", "c", "d"]:
        <tr class="$loop.parity">
            <td>$loop.index</td>
            <td>$c</td>
        </tr>
</table>
</body></html>
```

7) 进阶模板使用

1. server2.py

```
#-*-coding:utf-8-*-
import web
from web import form as form
urls = (
    '/', 'index',
    '/add', 'add'
)
render = web.template.render('template/')
class index:
    def GET(self):
        return render.post()
class add:
    def POST(self):
        print web.input()['title1']
        print web.data()
        raise web.seeother('/')
if __name__ == "__main__":
    app = web.application(urls, globals())
    app.run()
```

2. post.html

```
<html><head>
    <title>post test

    </title>
</head><body>
<form method="post" action="/add">
    <input type="text" name="title" />
    <input type="text" name="title1" />
    <input type="text" name="title2" />
    <input type="text" name="title3" />
    <input type="submit" value="Add" />
</form>

</body></html>
```

8) URL 控制

1. 问题：

如何为整个网站设计一个 URL 控制方案 / 调度模式

2. 解决：

`web.py` 的 URL 控制模式是简单的、强大的、灵活的。在每个应用的最顶部，你通常会看到整个 URL 调度模式被定义在元组中：

```
urls = (
    "/tasks/?", "signin",
    "/tasks/list", "listing",
    "/tasks/post", "post",
    "/tasks/chgpass", "chgpass",
    "/tasks/act", "actions",
    "/tasks/logout", "logout",
    "/tasks/signup", "signup"
)
```

这些元组的格式是：URL 路径，处理类 这组定义有多少可以定义多少。如果你并不知道 URL 路径和处理类之间的关系。

3. 路径匹配

你可以利用强大的正则表达式去设计更灵活的 URL 路径。比如 `/(test1|test2)` 可以捕捉 `/test1` 或 `/test2`。要理解这里的关键，匹配是依据 URL 路径的。比如下面的 URL：

```
http://localhost/myapp/greetings/hello?name=Joe
```

你可以捕捉 URL 的参数，然后用在处理类中：

```
/users/list/(.+), "list_users"
```

在 `list/` 后面的这块会被捕捉，然后作为参数被用在 GET 或 POST：

```
class list_users:
    def GET(self, name):
        return "Listing info about user: {}".format(name)
```

9) 一个完整的小例子

```
#-*-coding:utf-8-*-
import web

urls = (
    "/", "hello",
    "/chu888/?","chu888",
```

```
"/parmtest/(.+)", "listmy"
)
app = web.application(urls, globals())

class hello:
    def GET(self):
        return 'Hello, world!'
class chu888:
    def GET(self):
        return "你访问的是 chu888"
class listmy:
    def GET(self, name):
        return "Listing info about user: {}".format(name)
if __name__ == "__main__":
    app.run()
```

10) web.seeother 和 web.redirect 转向

1. 问题

在处理完用户输入后（比方说处理完一个表单），如何跳转到其他页面？

2. 解法

```
class SomePage:
    def POST(self):
        # Do some application logic here, and then:
        raise web.seeother('/someotherpage')
```

POST 方法接收到一个 post 并完成处理之后，它将给浏览器发送一个 303 消息和新网址。接下来，浏览器就会对这个新网址发出 GET 请求，从而完成跳转。

注意：web.seeother 和 web.redirect 不支持 0.3 以下版本。

3. 区别

用 web.redirect 方法似乎也能做同样的事情，但通常来说，这并太友好。因为 web.redirect 发送的是 301 消息—这是永久重定向。因为大多数 Web 浏览器会缓存新的重定向，所以当我们再次执行该操作时，会自动直接访问重定向的新网址。很多时候，这不是我们所想要的结果。所以在提交表单时，尽量使用 seeother。但是在下面要提到的这种场合，用 redirect 却是最恰当的：我们已经更改了网站的网址结构，但是仍想让用户书签/收藏夹中的旧网址不失效。

(注：要了解 seeother 和 redirect 的区别，最好是看一下 http 协议中不同消息码的含义。)

11) 包含应用

1. 实现

在 `blog.py` 中：

```
import web
urls = (
    "", "reblog",
    "/(.*)", "blog"
)

class reblog:
    def GET(self): raise web.seeother('/')

class blog:
    def GET(self, path):
        return "blog " + path

app_blog = web.application(urls, locals())
```

当前的主应用 `code.py`：

```
import web
import blog
urls = (
    "/blog", blog.app_blog,
    "/(.*)", "index"
)

class index:
    def GET(self, path):
        return "hello " + path

app = web.application(urls, locals())

if __name__ == "__main__":
    app.run()
```

12) 使用 XML

1. 问题

如何在 `web.py` 中提供 XML 访问？

如果需要为第三方应用收发数据，那么提供 `xml` 访问是很有必要的。

2. 解法

根据要访问的 `xml` 文件(如 `response.html`)创建一个 XML 模板。如果 XML 中有变量，就使用相应的模板标签进行替换。下面是一个例子：

```
$def with (code)
<?xml version="1.0"?>
<RequestNotification-Response>
<Status>$code</Status>
</RequestNotification-Response>
```

为了提供这个 XML，需要创建一个单独的 `web.py` 程序(如 `response.py`)，它要包含下面的代码。注意：要用 "`web.header('Content-Type', 'text/xml')`" 来告知客户端——正在发送的是一个 XML 文件。

```
import web

render = web.template.render('template/', cache=False)

urls = (
    '/(.*)', 'index'
)

app = web.application(urls, globals())

class index:
    def GET(self, code):
        web.header('Content-Type', 'text/xml')
        return render.index(code)

web.webapi.internalerror = web.debugerror
if __name__ == '__main__': app.run()
```

13) 获取 POST 数据

```
class RequestHandler():
    def POST():
        data = web.data() # 通过这个方法可以取到数据
```

1. login.html 模板

```
<form id="form1" name="form1" method="post" action="/post/">
    <h1>Sign-up form</h1>

    <p>This is the basic look of my form without table</p>
    <label>Name
        <span class="small">Add your name</span>
    </label>
    <input type="text" name="UserName" id="UserName"/>

    <label>Email
        <span class="small">Add a valid address</span>
    </label>
    <input type="text" name="Email" id="Email"/>

    <label>Password
        <span class="small">Min. size 6 chars</span>
    </label>
    <input type="text" name="PassWord" id="PassWord"/>
    <button type="submit">Sign-up</button>
    <div class="spacer"></div>
```

2. 主程序

```
-*-coding:utf-8-*-
import web

urls = (
    "/", "index",
    "/login/?","login",
    "/post/?","post"
)
render = web.template.render('template/')
app = web.application(urls, globals())
class index:
    def GET(self):
        return "Hello web.py!"
class login:
    def GET(self):
        return render.login()
class post:
    def POST(self):
        return web.input()['UserName']
if __name__ == "__main__":
    app.run()
```

14) 获取客户端信息

1. 问题

如何在代码中得到客户端信息？比如：来源页面(**referring page**)或是客户端浏览器类型

2. 解法

使用 `web.ctx` 即可。首先讲一点架构的东西：`web.ctx` 基于 `threadeddict` 类，又被叫做 `ThreadDict`。这个类创建了一个类似字典(**dictionary-like**)的对象，对象中的值都是与线程 `id` 相对应的。这样做很妙，因为很多用户同时访问系统时，这个字典对象能做到仅为某一特定的 HTTP 请求提供数据(因为没有数据共享，所以对象是线程安全的)

`web.ctx` 保存每个 HTTP 请求的特定信息，比如客户端环境变量。假设，我们想知道正在访问某页面的用户是从哪个网页跳转而来的：

3. 例子

```
class example:
    def GET(self):
        referer = web.ctx.env.get('HTTP_REFERER', 'http://google.com')
        raise web.seeother(referer)
```

上述代码用 `web.ctx.env` 获取 `HTTP_REFERER` 的值。如果 `HTTP_REFERER` 不存在，就会将 `google.com` 做为默认值。接下来，用户就会被重定向回到之前的来源页面。

`web.ctx` 另一个特性，是它可以被 `loadhook` 赋值。例如：当一个请求被处理时，会话(**Session**)就会被设置并保存在 `web.ctx` 中。由于 `web.ctx` 是线程安全的，所以我们可以象使用普通的 `python` 对象一样，来操作会话(**Session**)。

4. 'ctx'中的数据成员

Request

- `environ` 又被写做. `env` – 包含标准 WSGI 环境变量的字典。
- `home` – 应用的 http 根路径(译注：可以理解为应用的起始网址，协议+站点域名+应用所在路径)例：
`http://example.org/admin`
- `homedomain` – 应用所在站点(可以理解为协议+域名) `http://example.org`
- `homepath` – 当前应用所在的路径，例如： `/admin`
- `host` – 主机名(域名) + 用户请求的端口(如果没有的话，就是默认的 80 端口)，例如： `example.org,example.org:8080`
- `ip` – 用户的 IP 地址，例如： `xxx.xxx.xxx.xxx`
- `method` – 所用的 HTTP 方法，例如： `GET`
- `path` – 用户请求路径，它是基于当前应用的相对路径。在子应用中，匹配外部应用的那部分网址将被去掉。例如：主应用在 `code.py` 中，而子应用在 `admin.py` 中。在 `code.py` 中，我们将 `/admin` 关联到 `admin.app`。在 `admin.py` 中，将 `/stories` 关联到 `stories` 类。在 `stories` 中，`web.ctx.path` 就是 `/stories`，而非 `/admin/stories`。形如： `/articles/845`
- `protocol` – 所用协议，例如： `https`

- `query` – 跟在'?'字符后面的查询字符串。如果不存在查询参数，它就是一个空字符串。例如：`?fourlegs=good&twolegs=bad`
- `fullpath` 可以视为 `path + query` – 包含查询参数的请求路径，但不包括'homepath'。例如：`/articles/845?fourlegs=good&twolegs=bad`

Response

- `status` – HTTP 状态码（默认是'200 OK'）`401 Unauthorized` 未经授权
- `headers` – 包含 HTTP 头信息(headers)的二元组列表。
- `output` – 包含响应实体的字符串。

15) 文件上传

1. 简单的文件上传

```
#!/usr/bin/env python
# coding: utf-8
import web

class mytest:
    def GET(self):
        return "<b><h1>我扩展测试一下</h1><b>"

class Upload:
    def GET(self):
        return """<html><head></head><body>
<form method="POST" enctype="multipart/form-data" action="">
<input type="file" name="myfile" />
<br/>
<input type="submit" />
</form>
</body></html>"""

    def POST(self):
        x = web.input(myfile={})
        web.debug(x['myfile'].filename) # 这里是文件名
        web.debug(x['myfile'].value) # 这里是文件内容
        web.debug(x['myfile'].file.read()) # 或者使用一个文件对象
        print x['myfile'].filename
        print x['myfile'].value
        print x['myfile'].file.read()
        raise web.seeother('/todo/upload')
```

2. 基于新浪云的上传

```
#!/usr/bin/env python
```

```
# coding: utf-8
import web
import uuid
import random
import time
import sae.storage
from config import settings
from datetime import datetime

render = settings.render

class SaveUpload:
    def GET(self):
        # 初始化一个 Storage 客户端。
        s = sae.storage.Client()
        imagelist=s.list('pythondb')
        return render.SaveUpload(imagelist)

    # def GET(self):
    #     web.header("Content-Type","text/html; charset=utf-8")
    #     return """<html><head><title>文件上传并且保存 demo</title></head><body>
    #             <form method="POST" enctype="multipart/form-data" action="">
    #                 <input type="file" name="myfile" />
    #                 <br/>
    #                 <input type="submit" />
    #             </form>
    #         </body></html>"""

    def POST(self):
        x = web.input(myfile={})
        if 'myfile' in x:
            #读取所有文件内容
            imagevalue=x myfile file.read()
            # 初始化一个 Storage 客户端。
            s = sae.storage.Client()

            # 设置 object 的属性
            ob = sae.storage.Object(imagevalue)
            #保存文件
            saveuploadfile= str(time.time()) + str(random.random())+x myfile filename
            s.put('pythondb',saveuploadfile, ob)

raise web.seeother('/todo/saveupload')
```

```
$def with(imagelist)
<!DOCTYPE html>
<html>
```

```
<head>
    <title></title>
    <style type="text/css">
        .imagediv
        {
            width: 200px;
            height: 200px;
            border: solid 1px;
        }
    </style>
</head>
<body>
<form method="POST" enctype="multipart/form-data" action="">
    <input type="file" name="myfile" />
    <br/>
    <input type="submit" />
</form>
$for i in imagelist:
    <div></div>
</body>
</html>
```

十九、 Web.py 使用 Session

1) 问题

如何在 web.py 中使用 session

2) 解法

注意!!!: session 并不能在调试模式(Debug mode)下正常工作, 这是因为 session 与调试模式下的重调用相冲突(有点类似 firefox 下著名的 Firebug 插件, 使用 Firebug 插件分析网页时, 会在火狐浏览器之外单独对该网页发起请求, 所以相当于同时访问该网页两次), 下一节中我们会给出在调试模式下使用 session 的解决办法。
web.session 模块提供 session 支持。下面是一个简单的例子——统计有多少人正在使用 session(session 计数器):

```
import web
web.config.debug = False
urls = (
    "/count", "count",
    "/reset", "reset"
)
app = web.application(urls, locals())
```

```
session = web.session.Session(app, web.session.DiskStore('sessions'),
initializer={'count': 0})

class count:
    def GET(self):
        session.count += 1
        return str(session.count)

class reset:
    def GET(self):
        session.kill()
        return ""

if __name__ == "__main__":
    app.run()
```

`web.py` 在处理请求之前，就加载 `session` 对象及其数据；在请求处理完之后，会检查 `session` 数据是否被改动。如果被改动，就交由 `session` 对象保存。

上例中的 `initializer` 参数决定了 `session` 初始化的值，它是个可选参数。如果用数据库代替磁盘文件来存储 `session` 信息，只要用 `DBStore` 代替 `DiskStore` 即可。使用 `DBStore` 需要建立一个表，结构如下：

```
create table sessions (
    session_id char(128) UNIQUE NOT NULL,
    atime timestamp NOT NULL default current_timestamp,
    data text
);
```

`DBStore` 被创建要传入两个参数：`db` 对象和 `session` 的表名。

```
db = web.database(dbn='postgres', db='mydatabase', user='myname', pw='')
store = web.session.DBStore(db, 'sessions')
session = web.session.Session(app, store, initializer={'count': 0})
```

'`web.config`' 中的 `sessions_parameters` 保存着 `session` 的相关设置，`sessions_parameters` 本身是一个字典，可以对其修改。默认设置如下：

```
web.config.session_parameters['cookie_name'] = 'webpy_session_id'
web.config.session_parameters['cookie_domain'] = None
web.config.session_parameters['timeout'] = 86400, #24 * 60 * 60, # 24 hours in seconds
web.config.session_parameters['ignore_expiry'] = True
web.config.session_parameters['ignore_change_ip'] = True
web.config.session_parameters['secret_key'] = 'fLjUfxqXtfNoIldA0A0J'
web.config.session_parameters['expired_message'] = 'Session expired'
```

- `cookie_name` - 保存 `session id` 的 `Cookie` 的名称
- `cookie_domain` - 保存 `session id` 的 `Cookie` 的 `domain` 信息
- `timeout` - `session` 的有效时间，以秒为单位
- `ignore_expiry` - 如果为 `True`, `session` 就永不过期
- `ignore_change_ip` - 如果为 `true`, 就表明只有在访问该 `session` 的 IP 与创建该 `session` 的 IP 完全一致时，`session` 才被允许访问。
- `secret_key` - 密码种子，为 `session` 加密提供一个字符串种子
- `expired_message` - `session` 过期时显示的提示信息。

3) 在 template 下使用 Session

1. 问题：

我想在模板中使用 session (比如：读取并显示 `session.username`)

2. 解决：

在应用程序中的代码：

```
render = web.template.render('templates', globals={'context': session})
```

在模板中的代码：

```
<span>You are logged in as <b>$context.username</b></span>
```

你可以真正的使用任何符合语法的 python 变量名，比如上面用的 context。我更喜欢在应用中直接使用 'session'。

`server.py`

```
 #-*-coding:utf-8-*-
import web
web.config.debug = False
urls = (
    "/count", "count",
    "/reset", "reset",
    "/", "index"
)
app = web.application(urls, locals())

db = web.database(dbn='mysql', db='todo', user='root', pw='chu123')
store = web.session.DBStore(db, 'sessions')
session = web.session.Session(app, store, initializer={'count': 0})
render = web.template.render('template', globals={'context': session})

class index:
    def GET(self):
        return render.index()
class count:
    def GET(self):
        session.count += 1
        return str(session.count)

class reset:
    def GET(self):
        session.kill()
        return ""

if __name__ == "__main__":
    app.run()
```

index.html

```
<!DOCTYPE html>
<html>
<head>
    <title></title>
</head>
<body>
<h1>我是楚广明</h1>
<h1><span>You are logged in as <b>$context.count</b></span></h1>
</body>
</html>
```

4) 如何操作 Cookies

1. 问题

如何设置和获取用户的 Cookie?

2. 解法

对 web.py 而言，设置/获取 Cookie 非常方便。

3. 设置 Cookies

概述

```
setcookie(name, value, expires="", domain=None, secure=False):
```

- name (string) - Cookie 的名称，由浏览器保存并发送至服务器。
- value (string) -Cookie 的值，与 Cookie 的名称相对应。
- expires (int) - Cookie 的过期时间，这是个可选参数，它决定 cookie 有效时间是多久。以秒为单位。它必须是一个整数，而绝不能是字符串。
- domain (string) - Cookie 的有效域—在该域内 cookie 才是有效的。一般情况下，要在某站点内可用，该参数值该写做站点的域（比如.webpy.org），而不是站主的主机名（比如 wiki.webpy.org）
- secure (bool)- 如果为 True，要求该 Cookie 只能通过 HTTPS 传输。.

示例

用 web.setcookie() 设置 cookie，如下：

```
class CookieSet:
    def GET(self):
        i = web.input(age='25')
        web.setcookie('age', i.age, 3600)
        return "Age set in your cookie"
```

用 GET 方式调用上面的类将设置一个名为 age，默认值是 25 的 cookie(实际上，默认值 25 是在 web.input

中赋予 `i.age` 的，从而间接赋予 `cookie`，而不是在 `setcookie` 函数中直接赋予 `cookie` 的）。这个 `cookie` 将在一小时后(即 3600 秒)过期。

`web.setcookie()` 的第三个参数—"expires"是一个可选参数，它用来设定 `cookie` 过期的时间。如果是负数，`cookie` 将立刻过期。如果是正数，就表示 `cookie` 的有效时间是多久，以秒为单位。如果该参数为空，`cookie` 就永不过期。

4. 获得 Cookies

概述

获取 `Cookie` 的值有很多方法，它们的区别就在于找不到 `cookie` 时如何处理。

方法 1 (如果找不到 `cookie`, 就返回 `None`):

```
web.cookies().get(cookieName)  
#cookieName is the name of the cookie submitted by the browser
```

方法 2 (如果找不到 `cookie`, 就抛出 `AttributeError` 异常):

```
foo = web.cookies()  
foo.cookieName
```

方法 3 (如果找不到 `cookie`, 可以设置默认值来避免抛出异常):

```
foo = web.cookies(cookieName=defaultValue)  
foo.cookieName # return the value (which could be default)  
#cookieName is the name of the cookie submitted by the browser
```

示例:

用 `web.cookies()` 访问 `cookie`. 如果已经用 `web.setcookie()` 设置了 `Cookie`, 就可以象下面这样获得 `Cookie`:

```
class CookieGet:  
    def GET(self):  
        c = web.cookies(age="25")  
        return "Your age is: " + c.age
```

这个例子为 `cookie` 设置了默认值。这么做的原因是在访问时，若 `cookie` 不存在，`web.cookies()`就会抛出异常，如果事先设置了默认值就不会出现这种情况。

如果要确认 `cookie` 值是否存在，可以这样做:

```
class CookieGet:  
    def GET(self):  
        try:  
            return "Your age is: " + web.cookies().age  
        except:  
            # Do whatever handling you need to, etc. here.  
            return "Cookie does not exist."
```

或

```
class CookieGet:  
    def GET(self):  
        age=web.cookies().get('age')  
        if age:  
            return "Your age is: %s" % age  
        else:  
            return "Cookie does not exist."
```

5) 布局模板

1. 问题

如何让站点每个页面共享一个整站范围的模板？（在某些框架中，称为模板继承，比如 ASP.NET 中的母版页）

2. 方法

我们可以用 `base` 属性来实现：

```
render = web.template.render('templates/', base='layout')
```

现在如果你调用 `render.foo()` 方法，将会加载 `templates/foo.html` 模板，并且它将会被 `templates/layout.html` 模板包裹。

"`layout.html`" 是一个简单模板格式文件，它包含了一个模板变量，如下：

```
$def with (content)
<html>
<head>
    <title>Foo</title>
</head>
<body>
    $:content
</body>
</html>
```

在某些情况，如果不想使用基本模板，只需要创建一个没有 `base` 属性的 `reander` 对象，如下：

```
render_plain = web.template.render('templates/')
```

程序

```
-*-coding:utf-8-*-
import web
web.config.debug = True
urls = (
    "/","index"
)
app = web.application(urls, locals())
render = web.template.render('template/', base='layout')

class index:
    def GET(self):
        return render.index('楚广明 6666')

if __name__ == "__main__":
    app.run()
```

`layout.html`

```
$def with (content)
<html>
<head>
    <title>Foo</title>
```

```
</head>
<body>
$:content
</body>
</html>
```

index.html

```
$def with (name)
<h1>我是$name</h1>
```

3. Tip: 在布局文件 (layout.html) 中定义的页面标题变量

templates/index.html

```
$var title: This is title.

<h3>Hello, world</h3>
```

templates/layout.html

```
$def with (content)
<html>
<head>
    <title>$content.title</title>
</head>
<body>
$:content
</body>
</html>
```

4. Tip: 在其他模板中引用 css 文件，如下：

templates/login.html

```
$var cssfiles: static/login.css static/login2.css

hello, world.
```

templates/layout.html

```
$def with (content)
<html>
<head>
    <title>$content.title</title>

    $if content.cssfiles:
        $for f in content.cssfiles.split():
            <link rel="stylesheet" href="$f" type="text/css" media="screen" charset="utf-8"/>

</head>
<body>
```

```
$:content
</body>
</html>
```

输入的 HTML 代码如下：

```
<link rel="stylesheet" href="static/login.css" type="text/css" media="screen" charset="utf-8"/>
<link rel="stylesheet" href="static/login2.css" type="text/css" media="screen"
charset="utf-8"/>
```

二十、 Web.py 支持的语法模板

1) 使用 Templetor 模板语法

1. Introduction

web.py 的模板语言叫做 Templetor，它能负责将 python 的强大功能传递给模板系统。在模板中没有重新设计语法，它是类 python 的。如果你会 python，你可以顺手拈来。

这是一个模板示例：

```
$def with (name)
Hello $name!
```

第一行表示模板定义了一个变量 name。第二行中的 \$name 将会用 name 的值来替换。

2. 使用模板系统

通用渲染模板的方法：

```
render = web.template.render('templates')
return render.hello('world')
```

render 方法从模板根目录查找模板文件，render.hello(..) 表示渲染 hello.html 模板。实际上，系统会在根目录去查找叫 hello 的所有文件，直到找到匹配的。（事实上他只支持 .html 和 .xml 两种）

完整程序：

```
-*-coding:utf-8-*-
import web
web.config.debug = True
urls = (
    "/","index"
)
app = web.application(urls, locals())
render = web.template.render('template', globals())

class index:
    def GET(self):
        return render.index('楚广明 6666')
```

```
if __name__ == "__main__":
    app.run()
```

模板：

```
$def with (name)
<!DOCTYPE html>
<html>
<head>
    <title></title>
</head>
<body>
<h1>我是$name</h1>
</body>
</html>
```

除了上面的使用方式，你也可以直接用文件的方式来处理模板 frender：

```
hello = web.template.frender('templates/hello.html')
render hello('world')
```

直接使用字符串方式：

```
template = "$def with (name)\nHello $name"
hello = web.template.Template(template)
return hello('world')
```

3. 表达式用法

特殊字符 \$ 被用于特殊的 python 表达式。表达式能够被用于一些确定的组合当中 () 和 {}：

```
Look, a $string.
Hark, an ${arbitrary + expression}.
Gawk, a ${dictionary[key].function('argument')}.
Cool, a ${(limit)ing.}
```

4. 赋值

有时你可能需要定义一个新变量或给一些变量重新赋值，如下：

```
$ bug = get_bug(id)
<h1>$bug.title</h1>
<div>
    $bug.description
<div>
```

注意 \$ 在赋值变量名称之前要有一个空格，这有区别于常规的赋值用法。

完整程序：

```
#-*-coding:utf-8-*-
import web
web.config.debug = True
```

```
urls = (
    "/","index"
)
app = web.application(urls, locals())
render = web.template.render('template', globals)

class index:
    def GET(self):
        return render.index('楚广明 6666')

if __name__ == "__main__":
    app.run()
```

```
$def with (name)
<!DOCTYPE html>
<html>
<head>
    <title></title>
</head>
<body>
<h1>我是$name</h1>
$def get_bug(id='1'):
    Hello $name!
$ bug = get_bug(id)
<h1>$bug</h1>
</body>
</html>
```

5. 过滤

模板默认会使用 `web.websafe` 过滤 `html` 内容(`encoding` 处理)。

```
>>> render.hello("1 < 2")
"Hello 1 &lt; 2"
```

不需要过滤可以在 `$` 之后 使用 `:`。示例：该 `html` 内容不会被转义

```
$:form.render()
```

6. 新起一行用法

在行末添加 \ 代表显示层该内容不会被真实处理成一行。

```
If you put a backslash \
at the end of a line \
(like these) \
then there will be no newline.
```

7. 转义 \$

使用 \$\$ 可以在输出的时候显示字符 \$.

```
Can you lend me $$50?
```

8. 注释

\$# 是注释指示符。任何以 \$# 开始的某行内容都被当做注释。

```
$# this is a comment
Hello $name.title()! $# display the name in title case
```

9. 控制结构

模板系统支持 **for**, **while**, **if**, **elif** 和 **else**。像 python 一样，这里是需要缩进的。

```
$for i in range(10):
    I like $i

$for i in range(10): I like $i
$ a= [1,2,3,4,5]
$while a:
    hello $a.pop()
$ times=100
$ max=200
$if times > max:
    Stop! In the name of love.
$else:
    Keep on, you can do it.
```

for 循环内的成员变量只在循环内发生可用：

```
loop.index: the iteration of the loop (1-indexed)
loop.index0: the iteration of the loop (0-indexed)
loop.first: True if first iteration
loop.last: True if last iteration
loop.odd: True if an odd iteration
loop.even: True if an even iteration
loop.parity: "odd" or "even" depending on which is true
loop.parent: the loop above this in nested loops
```

有时候，他们使用起来很方便：

```
<table>
$for c in ["a", "b", "c", "d"]:
    <tr class="$loop.parity">
        <td>$loop.index</td>
        <td>$c</td>
    </tr>
</table>
```

10. 使用 def

可以使用 \$def 定义一个新的模板函数，支持使用参数。

```
$def say_hello(name='world'):
    Hello $name!

$say_hello('web.py')
$say_hello()
```

其他示例：

```
$def tr(values):
    <tr>
        $for v in values:
            <td>$v</td>
    </tr>

$def table(rows):
    <table>
        $for row in rows:
            $:row
    </table>

$ data = [['a', 'b', 'c'], [1, 2, 3], [2, 4, 6], [3, 6, 9] ]
$:table([tr(d) for d in data])
```

代码

可以在 code 块书写任何 python 代码：

```
$code:
    x = "you can write any python code here"
    y = x.title()
    z = len(x + y)

    def limit(s, width=10):
        """limits a string to the given width"""
        if len(s) >= width:
            return s[:width] + "..."
        else:
            return s
```

And we are back to template.

The variables defined in the code block can be used here.

For example, \$limit(x)

11. 使用 var

var 块可以用来定义模板结果的额外属性：

```
$def with (title, body)
```

```
$var title: $title
$var content_type: text/html

<div id="body">
$body
</div>
```

以上模板内容的输出结果如下：

```
>>> out = render.page('hello', 'hello world')
>>> out.title
u'hello'
>>> out.content_type
u'text/html'
>>> str(out)
'\n\n<div>\nhello world\n</div>\n'
```

12. 内置和全局

像 `python` 的任何函数一样，模板系统同样可以使用内置以及局部参数。很多内置的公共方法像 `range`, `min`, `max` 等，以及布尔值 `True` 和 `False`，在模板中都是可用的。部分内置和全局对象也可以使用在模板中。
全局对象可以使用参数方式传给模板，使用 `web.template.render`:

```
import web
import markdown

globals = {'markdown': markdown.markdown}
render = web.template.render('templates', globals=globals)
```

内置方法是否可以在模板中也是可以被控制的：

```
# 禁用所有内置方法
render = web.template.render('templates', builtins={})
```

这个比较难理解，我可以给出一个例子：

模板很简单 `index.html`

```
$:title
```

`code.py`

```
-*- coding:utf-8 -*-
import web
import sys
#如果页面有修改可以即时体现
reload(sys)
sys.setdefaultencoding('utf-8')
#定义 url, 将地址映射到对应的类
urls = (
    "/", "index",
)
app = web.application(urls, globals())

def render(params={}, partial=False):
```

```
global_vars = dict(params.items())

if partial:
    #print "true"
    return web.template.render('templates/', globals=global_vars)
else:
    #print "false"
    return web.template.render('templates/', base='layout', globals=global_vars)

#定义 index 类
class index:
    #get 请求
    def GET(self):
        return render({'title':'这是一个测试'},True).index()

if __name__ == "__main__":
    app.run()
```

2) 使用 Jinja2 模板语法

1. 上手的一个小例子

```
 #-*- coding:utf-8 -*-
import web
# 引入 web.py 对调用 jinja2 的模块
from web.contrib.template import render_jinja


urls = (
    "/", "index",
)
app = web.application(urls, globals())


render = render_jinja(
    'templates', # 模板存放的目录名称
    encoding='utf-8', # 模板使用的编码
)


class index:

    def GET(self):
        return render.index(word="Hello World") # 使用模板目录下的 index.html 模板

if __name__ == "__main__":
```

```
app.run()
```

```
<!DOCTYPE html>
<html>
<head>
    <title></title>
</head>
<body>
    <h1>{{ word }}</h1>
</body>
</html>
```

二十一、 Web.py 中使用数据库

1) 使用 web.py 自带的 DB 封装库

首先需要大家用 `easy_install` 安装 `mysql` 的驱动，也可以自行去下载软件。有一定要注意就是 **64** 位的驱动与 **32** 位的是不兼容的，这一点要注意了。

`web.py` 是一个非常精巧的 `web` 框架，不过其自带的 `db` 模块也是非常精简而高效。

1. 一个简单的数据库访问测试

```
 #-*-coding:utf-8-*-
import sys
import web
#定义数据库对象 db
db = web.database(dbn='mysql', db='todolist', user='root', pw='chu123')
#定义要操作的数据库名称
tb = 'todo'
def get_by_id(id):
    s = db.select(tb, where='id=$id', vars=locals())
    if not s:
        return False
    return s[0]
if __name__=="__main__":
    resultValue=get_by_id(4)
    print resultValue
```

如果语句里面使用了\$符号，就要添加 `vars=locals()` 这一句，不然可能会报错。里面的\$id 是传递过来的 id 变量。

2. 一个复杂一点的例子

```
#-*-coding:utf-8-*-
import sys
import web
import time
#定义数据库对象 db
db = web.database(dbn='mysql', db='todolist', user='root', pw='chu123')
#定义要操作的数据库名称
tb = 'todo'
def getTableInfo(id=1):
    result = db.query('select * from todo where id>$id', vars=locals())
    for s in result:
        print s.id, s.title
if __name__=="__main__":
    #简单查询
    result = db.query('select * from todo where id>$id', vars={'id':1})
    for s in result:
        print s.id, s.title
    #插入,第一参数是表名 todo
    db.insert('todo',
title='Michael',finished=0,post_date=time.strftime('%Y-%m-%d',time.localtime(time.time())))
    getTableInfo(5)
    #更新
    db.update('todo', where='id=$id', vars={'id':5}, title='楚广明')
    getTableInfo(5)
    #删除
    db.delete('todo', where='id=$id', vars={'id':6})
    getTableInfo(5)
```

3. 更加复杂的例子

```
#-*-coding:utf-8-*-
import sys
import web
import time
#定义数据库对象 db
db = web.database(dbn='mysql', db='todolist', user='root', pw='chu123')
#定义要操作的数据库名称
tb = 'todo'
if __name__=="__main__":
    # 查询表
    entries = db.select('todo')
```

```
# where 条件
myvar = dict(title="楚广明")
results = db.select('todo', myvar, where="title = $title")
for s in results:
    print s.id
results = db.select('todo', where="id>100")
for s in results:
    print s.id
# 查询具体列
results = db.select('todo', what="id,title")
# order by
results = db.select('todo', order="post_date DESC")
# group
results = db.select('todo', group="title")
# limit
results = db.select('todo', limit=10)
#update
db.update('todo', where="id = 10", title = "foo")
#delete
db.delete('todo', where="id=10")
#复杂
# count
results = db.query("SELECT COUNT(*) AS total_users FROM todo")
print results[0].total_users

# join
#results = db.query("SELECT * FROM entries JOIN users WHERE entries.author_id =
users.id")

# 防止 SQL 注入可以这么干
#results = db.query("SELECT * FROM users WHERE id=$id", vars={'id':10})

'''多数据库操作
db1 = web.database(dbn='mysql', db='dbname1', user='foo')
db2 = web.database(dbn='mysql', db='dbname2', user='foo')

print db1.select('foo', where='id=1')
print db2.select('bar', where='id=5')
'''

'''事务
t = db.transaction()
try:
    db.insert('person', name='foo')
    db.insert('person', name='bar')
except:
    t.rollback()
    raise
```

```
else:  
    t.commit()  
  
# Python 2.5+ 可以用 with  
from __future__ import with_statement  
with db.transaction():  
    db.insert('person', name='foo')  
    db.insert('person', name='bar')  
...  
...
```

二十二、 使用 SQLAlchemy

1) 简介

web 开发中一般的主要版块就是视图、模板、数据库操作，前面已经讲了最基础的视图跟模板，这一章主要是介绍下数据库操作，如第一章前言所说，我们的数据库操作是使用 `sqlalchemy` 来实现的。

对象关系映射器（Object Relational Mappers, ORM）在过去数年吸引了不少人的目光。主要原因是 ORM 经常会在 Web 应用程序框架中被提起，因为它是快速开发（Rapid Development）栈中的关键组件。Django 和 Ruby on Rails 等 Web 框架采用了设计一个独立栈的方法，将自主开发的 ORM 紧密集成到该框架中。而其他框架，如 Pylons、Turbogears 和 Grok，则采用更加基于组件的架构结合可交换的第三方组件。两种方法都有各自的优势：紧密集成允许非常连贯的体验（如果问题映射到框架），而基于组件的架构则允许最大的设计灵活性。但是，本文的主题并不是 Web 框架；而是 SQLAlchemy。

SQLAlchemy 在构建在 WSGI 规范上的下一代 Python Web 框架中得到了广泛应用，它是由 Mike Bayer 和他的核心开发人员团队开发的一个单独的项目。使用 ORM 等独立 SQLAlchemy 的一个优势就是它允许开发人员首先考虑数据模型，并能决定稍后可视化数据的方式（采用命令行工具、Web 框架还是 GUI 框架）。这与先决定使用 Web 框架或 GUI 框架，然后再决定如何在框架允许的范围内使用数据模型的开发方法极为不同。

不同于很多 web 开发，python 开发 web 的数据库很多是通过建立数据表类，绑定数据库连接，然后自动创建数据表，这种设计很大程度上降低了在开发阶段修改数据库结构的麻烦，而且由于表是对应一个个类创建而成，可以很清晰的在任何时候了解到该 app 的数据库是如何设计的，这对维护的意义也是非常大的。

什么是 WSGI？

WSGI 是下一代 Python Web 框架、应用程序和服务器应该遵循的规范。WSGI 中一个有趣的方面是创建 Python 中间件，并在使用 Python 或任何语言创建的 Web 应用程序中使用。请参阅 参考资料，获取关于 WSGI 和 WSGI 社区（Pyperfitters）的大量链接。

SQLAlchemy 的一个目标是提供能兼容众多数据库（如 SQLite、MySQL、Postgres、Oracle、MS-SQL、SQLServer 和 Firebird）的企业级持久性模型。SQLAlchemy 正处于积极开发阶段，当前最新的 API 将围绕版本 0.5 设计。请参阅参考资料部分，获取官方 API 文档、教程和 SQLAlchemy 书籍的链接。

SQLAlchemy 取得成功的一个证明就是围绕它已建立了丰富的社区。针对 SQLAlchemy 的扩展和插件包括：declarative、Migrate、Elixir、SQLSoup、django-sqlalchemy、DBSprockets、FormAlchemy 和 z3c.sqlalchemy。在本文中，我们将学习一篇关于新 0.5 API 的教程，探究一些第三方库，以及如何在 Pylons 中使用它们。

谁是 Mike Bayer？

Michael Bayer 是居住在纽约的一名软件承包商，他拥有十余年处理各类关系数据库的经验。他曾使用 C、Java™ 和 Perl 编写了许多自主研发的数据库抽象层，并在 Major League Baseball 与大量多服务器 Oracle 系统打了多年交道，借助这些经验，他成功编写了“终极工具包” SQLAlchemy，用于生成 SQL 和处理数据库。

例子一

```
# -*- coding:utf-8 -*-
from sqlalchemy import *
from sqlalchemy.ext.declarative import declarative_base
from sqlalchemy.orm import scoped_session, sessionmaker
db_config = {
    'host': 'localhost',
    'user': 'root',
    'passwd': 'sdss',
    'db': 'test',
    'charset': 'utf8'
}

engine = create_engine('mysql://{}:{}@{}:{}/{}?charset={}'.format(
    db_config['user'],
    db_config['passwd'],
    db_config['host'],
    db_config['db'],
    db_config['charset']), echo=True)

metadata = MetaData()
users_table = Table('users222', metadata,
    Column('id', Integer, primary_key=True),
    Column('name', String(50)),
    Column('fullname', String(50)),
    Column('password', String(100))
)
metadata.create_all(engine)
```

例子二

```
# -*- coding:utf-8 -*-
from sqlalchemy import *
from sqlalchemy.ext.declarative import declarative_base
from sqlalchemy.orm import scoped_session, sessionmaker

# 定义数据库的账号、端口、密码、数据库名，使用的连接模块，这里用的是 mysql+mysqldb
engine = create_engine(
    'mysql+mysqldb://root:dddd123@localhost:3306/todo?charset=utf8',
    echo=True # 是否输出数据库操作过程，很方便调试
)
Base = declarative_base()

class User(Base):
    __tablename__ = "user2222"
    id = Column(Integer, primary_key=True)
    name = Column(String(20), unique=True)
    email = Column(String(32), unique=True)
    password = Column(String(32))
```

```
superuser = Column(Boolean, default=False)

metadata = Base.metadata

if __name__ == "__main__":
    metadata.create_all(engine) #运行 python models.py 就会自动创建定义的所有表
```

以上就是 User 表的定义，可以在 User 类里非常清晰的看出这个数据表是如何设计的。接下来先在 mysql 创建一个叫"tech"的数据库，接着只要在命令行运行 python models.py 就可以看到数据库中出现了 User 表了，用数据库操作工具看看是否与定义的内容一样吧：）

2) 初始化 Sqlite 数据库

```
from sqlalchemy import create_engine
from sqlalchemy import Column, Integer, String
from sqlalchemy.ext.declarative import declarative_base

engine = create_engine('sqlite:///mydatabase.db', echo=True)
Base = declarative_base()

class User(Base):
    __tablename__ = 'users'

    id = Column(Integer, primary_key=True)
    name = Column(String)
    fullname = Column(String)
    password = Column(String)

    def __init__(self, name, fullname, password):
        self.name = name
        self.fullname = fullname
        self.password = password

    def __repr__(self):
        return "<User('%s', '%s', '%s')>" % (self.name, self.fullname, self.password)

users_table = User.__table__
metadata = Base.metadata
if __name__ == "__main__":
    metadata.create_all(engine)
```

3) 插入数据库

```
# -*- coding:utf-8 -*-
from sqlalchemy import *
from sqlalchemy.ext.declarative import declarative_base
from sqlalchemy.orm import scoped_session, sessionmaker

# 定义数据库的账号、端口、密码、数据库名，使用的连接模块，这里用的是 mysql
db_config = {
    'host': 'localhost',
    'user': 'root',
    'passwd': 'sdss',
    'db': 'todo',
    'charset': 'utf8'
}

engine = create_engine('mysql://{}:{}@{}:{}/{}?charset=%s'.format(
    db_config['user'],
    db_config['passwd'],
    db_config['host'],
    db_config['db'],
    db_config['charset']),
    echo=True)

# 定义一个函数，用来获取 sqlalchemy 的 session
def bindSQL():
    return scoped_session(sessionmaker(bind=engine))

Base = declarative_base()

class User(Base):
    __tablename__ = "user22"
    id = Column(Integer, primary_key=True)
    name = Column(String(20), unique=True)
    email = Column(String(32), unique=True)
    password = Column(String(32))
    superuser = Column(Boolean, default=False)
    def __init__(self, name='', email='', password='', superuser=True):
        self.name = name
        self.email = email
        self.password = password
        self.superuser = superuser

metadata = Base.metadata
```

```

if __name__ == "__main__":
    metadata.create_all(engine) # 运行 python models.py 就会自动创建定义的所有表
    my_vp = User('楚广明', 'ss3林@sss.com', '22月333', False)
    session=bindSQL()
    session.add(my_vp)
    session.commit()

```

除了将 `user` 表创建出来，我们并没有进行其他的操作，而往往我们的 `web` 开发，总需要在网站中设计一个超级管理员，这个超级管理员不是注册出来的，而是在数据库创建初期就有的，那么，我们可以在 `models.py` 中创建 `user` 表，并同时向 `user` 表插入超级管理员的信息，而在插入超级管理员或者以后的修改密码、新增用户的时候，我们的密码都是需要加密的，那么这时就可以添加一个事件触发，当插入设置 `password` 字段时，自动加密，好，需求有了，接下来就是行动了，我们把原来的 `models.py` 修改成以下代码：

```

# -*- coding:utf-8 -*-
from sqlalchemy import *
from sqlalchemy import event
from sqlalchemy.ext.declarative import declarative_base
from sqlalchemy.orm import scoped_session, sessionmaker
import hashlib

# 这里定义一个 password 加密混淆
password_prefix = "Ad%cvcsadefr^!deaf"

# 定义数据库的账号、端口、密码、数据库名，使用的连接模块，
# 这里用的是 mysql
engine = create_engine(
    'mysql+mysqldb://root:dddd123@localhost:3306/todo?charset=utf8',
    echo=True # 是否输出数据库操作过程，很方便调试
)

# 定义一个函数，用来获取 sqlalchemy 的 session

def bindSQL():
    return scoped_session(sessionmaker(bind=engine))

Base = declarative_base()
# 定义数据表使用 InnoDB
Base.__table_args__ = {'mysql_engine': 'InnoDB'}


class User(Base):
    __tablename__ = "user5555"
    id = Column(Integer, primary_key=True)
    name = Column(String(20), unique=True)
    email = Column(String(32), unique=True)

```

```
password = Column(String(32))
superuser = Column(Boolean, default=False)

metadata = Base.metadata

# 定义一个回调函数用于响应触发事件

def setPassword(target, value, oldvalue, initiator):
    # 如果新设置的值与原有的值相等，那么说明用户并没有修改密码，返回原先的值
    if value == oldvalue:
        return oldvalue
    # 如果新值与旧值不同，说明密码发生改变，进行加密，加密方法可以根据自己需求改变
    return hashlib.md5("%s%s" % (password_prefix, value)).hexdigest()

# 设置事件监听，event.listen(表单或表单字段，触发事件，回调函数，是否改变插入值)
event.listen(User.password, "set", setPassword, retval=True)

# 为了避免重复插入数据，定义一个 get_or_create 函数
def get_or_create(session, model, **kwargs):
    if "defaults" in kwargs:
        defaults = kwargs["defaults"]
        del kwargs["defaults"]
    else:
        defaults = {}

    instance = session.query(model).filter_by(**kwargs).first()
    if instance:
        return instance, False
    else:
        kwargs.update(defaults)
        instance = model(**kwargs)
        session.add(instance)
        session.flush()
        session.refresh(instance)
        return instance, True

# 定义初始化函数

def initModel():
    metadata.create_all(engine) # 创建数据库
    db = bindSQL() # 获取 sqlalchemy 的 session
    # 创建超级管理员，这里为了避免多次运行 initModel
    # 而发生重复插入的情况，使用了 get_or_create 方法
    obj, created = get_or_create(
        db,
        User,
        name="administrator",
```

```
defaults={  
    "email": "332535694@qq.com",  
    "password": "administrator",  
    "superuser": True  
}  
)  
db.commit() # 记得 commit 哟，不然数据最后还是没插入  
db.remove()  
  
if __name__ == "__main__":  
    initModel()
```

4) 在视图中使用 sqlalchemy

1. 首先创建数据库 models.py

```
from sqlalchemy import create_engine  
from sqlalchemy import Column, Integer, String  
from sqlalchemy.ext.declarative import declarative_base  
  
engine = create_engine('sqlite:///mydatabase.db', echo=True)  
Base = declarative_base()  
  
  
class User(Base):  
    __tablename__ = 'users'  
  
    id = Column(Integer, primary_key=True)  
    name = Column(String)  
    fullname = Column(String)  
    password = Column(String)  
  
    def __init__(self, name, fullname, password):  
        self.name = name  
        self.fullname = fullname  
        self.password = password  
  
    def __repr__(self):  
        return "<User('%s', '%s', '%s')>" % (self.name, self.fullname, self.password)  
  
  
users_table = User.__table__  
metadata = Base.metadata  
if __name__ == "__main__":  
    metadata.create_all(engine)
```

2. 使用 app.py

```
import string
import random
import web

from sqlalchemy.orm import scoped_session, sessionmaker
from models import *

urls = (
    "/", "add",
    "/view", "view"
)

def load_sqla(handler):
    web.ctx.orm = scoped_session(sessionmaker(bind=engine))
    try:
        return handler()
    except web.HTTPError:
        web.ctx.orm.commit()
        raise
    except:
        web.ctx.orm.rollback()
        raise
    finally:
        web.ctx.orm.commit()

app = web.application(urls, globals())
app.add_processor(load_sqla)

class add:
    def GET(self):
        web.header('Content-type', 'text/html')
        fname = "".join(random.choice(string.letters) for i in range(4))
        lname = "".join(random.choice(string.letters) for i in range(7))
        u = User(name=fname, fullname=fname + ' ' + lname, password=542)
        web.ctx.orm.add(u)
        return "added:" + web.websafe(str(u)) \
            + "<br/>" \
            + '<a href="/view">view all</a>'

class view:
```

```
def GET(self):
    web.header('Content-type', 'text/plain')
    return "\n".join(map(str, web.ctx.orm.query(User).all()))

if __name__ == "__main__":
    app.run()
```

二十三、Json 的使用

首先 Python 内置的 json 包是支持 json 的解析的，我们平时如果不涉及复杂的业务，这个内置的包功能完全够用。先上一个例子

```
#!/usr/bin/python
#coding:utf-8
import json
#Function:Analyze json script
#Json is a script can descript data structure as xml,
#for detail, please refer to "http://json.org/json-zh.html".

#Note:
#1.Also, if you write json script from python,
#you should use dump instead of load. pleaser refer to "help(json)".

#json file:
#The file content of temp.json is:
#{{
# "name":"00_sample_case1",
# "description":"an example."
#}}
#f = file("temp.json");
#s = json.load(f)
#print s
#f.close

#json string:
s = json.loads('{"name":"test", "type":{"name":"seq", "parameter":["1", "2"]}}')
print s
print s.keys()
print s["name"]
print s["type"]["name"]
print s["type"]["parameter"][1]
```

二十四、 urllib 模块

1) urlopen 的基本使用

`urllib.urlopen(url[, data[, proxies]])` :
url: 表示远程数据的路径
data: 以 post 方式提交到 url 的数据
proxies: 用于设置代理

`urllib.urlopen(url[, date[, poxies]])` 这个方法打开一个 URL 并返回远程 url 的类文件对象，我们可以像操作本地文件一样操作 url 类的文件对象获取数据，其中参数 url 是远程数据的地址（URL），date 是以 post 方式提交到服务器的数据，poxies 是设置代理端口的；

`urlopen` 返回的类文件对象有以下方法：

- 1 `read()`, `readline()`, `readlines()` 这些方法跟操作文件一样；
- 2 `fileno()` 以整数返回文件描述符；
- 3 `close()` 关闭链接；
- 4 `getcode()` 返回 HTTP 响应码，例如：成功会返回 200，未找到文件的返回 404；
- 5 `geturl()` 用于返回请求的 URL；

```
#coding:utf-8
import urllib
import sys
google=urllib.urlopen("http://www.baidu.com")
#获取服务器的表头信息
print "this is a header:\n%s"%google.info()
print "-----"
#返回整数状态码
print "this is a status:\n%s"%google.getcode()
print "-----"
#返回 url
print "this is a url:\n%s"%google.geturl()
print "-----"
#返回文件描述信息
print google.read().decode('utf-8')
#print google.read()
#获取系统默认编码
#print sys.setdefaultencoding()
```

2) urlretrieve 方法的使用

```
urllib.urlretrieve(url[, filename[, reporthook[, data]]]):  
filename 指定保存到本地的路径（若未指定该，urllib 生成一个临时文件保存数据）  
reporthook 回调函数，当连接上服务器、以及相应的数据块传输完毕的时候会触发该回调
```

data 指 post 到服务器的数据

urllib.urlretrieve(url[, filename[, reporthook[, date]]]) 此函数将远程的数据下载到本地，其中参数 url 是 URL 字符串，filename 指定数据保存到本地的路径；reporthook(block, size, total)是一个回调函数当链接上服务器，以及相应的数据传送完毕时会触发该函数（即每下载一块(block)就调用一次回调函数），我们可以用该函数来显示当前数据下载的进度，也可以用来限速，date 是指 post 到服务器的数据；该方法返回一个包含两个元素的元组(filename, headers)，filename 表示保存到本地的路径，header 表示服务器的响应头。

```
from __future__ import unicode_literals  
import urllib  
import re  
  
def callback_f(downloaded_size, block_size, remote_total_size):  
    per = 100.0 * downloaded_size * block_size / remote_total_size  
    if per > 100:  
        per = 100  
    print "%.2f%%" % per  
  
def getHtml(url):  
    page=urllib.urlopen(url)  
    html=page.read()  
    page.close()  
    return html  
  
def getImg(html):  
    imgre=re.compile(r"""<img\s.*?\s?src\s*=\s*['|"]?([^\s']+).*?>""",re.I)  
    imglist=imgre.findall(html)  
    i=0  
    for imgurl in imglist:  
        if not imgurl.find("http://"):  
            urllib.urlretrieve(imgurl,"img/%d.jpg"%(i),callback_f)  
            i+=1  
            print imgurl  
ht=getHtml(r"http://www.cnbeta.com/articles/234303.htm")  
getImg(ht)
```

3) url 编码

- 1 urllib.quote(str[, safe]) 对字符串进行编码，而参数 safe 指定不需要编码的字符；
- 2 urllib.unquote(str) 对字符串进行解码；
- 3 urllib.quote_plus(str[, safe]) 和 quote 类似，只是把参数 str 中的空格转化成“+” 而 quote 是把

空格转化成“%20”

- 4 `urllib.unquote_plus(str)` 对字符串进行解码；
- 5 `urllib.urlencode(query[,dosep])` 参数 `query` 可以是两个元素的元组也可以是一个字，将其转成 `url`；
- 6 `urllib.url2pathname(path)` 将 `url` 路径转化成本地路径；
- 7 `urllib.pathname2url` 将本地路径转化成 `url` 路径；

```
#coding:utf-8
import urllib
date="whoiam=i am student+i like python"
#对字符串编码
date1=urllib.quote(date)
print date1
#对字符串解码
print urllib.unquote(date1)
print urllib.unquote_plus(date1)
#将参数试点转成 url
date5=urllib.urlencode({"name":"python","love":"python"})
print date5
```

4) 使用 `httpclient` 抓取

```
httpclient.HTTPConnection ( host [ , port [ ,strict [ , timeout ]]] )  
host 表示服务器主机  
port 为端口号，默认值为 80  
strict 的 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  
可选参数 timeout 表示超时时间。
```

`HTTPConnection` 提供的方法：

```
- HTTPConnection.request ( method , url [ ,body [ , headers ]])  
调用 request 方法会向服务器发送一次请求  
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。
```

```
- 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。
```

```
- HTTPConnection.connect ()  
连接到 Http 服务器。
```

```
- HTTPConnection.close ()
```

关闭与服务器的连接。

- `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。

`httpplib.HTTPResponse`

-`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：

-`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。

-`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。

-`HTTPResponse.getheaders()`

以列表的形式返回所有的头信息。

-`HTTPResponse.msg`

获取所有的响应头信息。

-`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。

-`HTTPResponse.status`

获取响应的状态码。如：200 表示请求成功。

-`HTTPResponse.reason`

返回服务器处理请求的结果说明。一般为“OK”

```
#!/usr/bin/python
# -*- coding:utf-8 -*-

def use_httpplib():
    import httpplib
    conn = httpplib.HTTPConnection("www.baidu.com")
    i_headers = {"User-Agent": "Mozilla/5.0 (Windows; U; Windows NT 5.1; zh-CN; rv:1.9.1) Gecko/20090624 Firefox/3.5",
                 "Accept": "text/plain"}
    conn.request("GET", "/", headers=i_headers)
    r1 = conn.getresponse()
    print "version:", r1.version
    print "-----"
    print "reason:", r1.reason
```

```
print "-----"
print "status:", r1.status
print "-----"
print "msg:", r1.msg
print "-----"
print "headers:", r1.getheaders()
print "-----"
data = r1.read()
print data.decode("utf-8")
print "-----"
print "data length:%s"%len(data)
conn.close()

if __name__ == "__main__":
    use_httplib()
```

二十五、 urllib2 的使用

1) 最简单的爬虫

网络爬虫是一个自动提取网页的程序，它为搜索引擎从万维网上下载网页，是搜索引擎的重要组成。`python` 的 `urllib\urllib2` 等模块很容易实现这一功能，下面的例子实现的是对 `baidu` 首页的下载。具体代码如下：

```
import urllib2
page=urllib2.urlopen("http://www.baidu.com")
print page.read().decode('utf-8')
```

2) 提交表单数据

1. 用 GET 方法提交数据

提交表单的 `GET` 方法是把表单数据编码至 `URL`。在给出请示的页面后，加上问号，接着是表单的元素。

如在百度中搜索“马伊琍”得到 `url` 为：

http://www.baidu.com/s?wd=%E9%A9%AC%E4%BC%8A%E7%90%8D&pn=100&rn=20&ie=utf-8&usm=4&rsv_page=1

其中？后面为表单元素。`wd=%E9%A9%AC%E4%BC%8A%E7%90%8D` 表示搜索的词是“马伊琍”，`pn` 表示从第 100 条信息所在页开始显示（感觉是这样，我试了几次，当写 100 时，从其所在页显示，但如果写 10，就是从第 1 页显示），`rn=20` 表示每页显示 20 条，`ie=utf-8` 表示编码格式，`usm=4` 没明白是什么意思，换了 1、2、3 试了下，没发现什么变化，`rsv_page=1` 表示第几页。如果要下载以上页面比较简单的方法是直接用上面的网址进行提取。如代码：

```
#coding:utf-8
import urllib2
```

```
keyword=urllib2.quote('马伊琍')
page=urllib2.urlopen("http://www.baidu.com/s?wd="+keyword+"&pn=100&rn=20&ie=utf-8&usm=4
&rsv_page=1")
print page.read()
```

二十六、 操作 Memcached

简单介绍：memcached 很强大，它可以支持分布式的共享内存缓存，大型站点都用它。对小站点来说，有足够的内存的话，使用它也可以得到超赞的效果。

1) 安装

1. Linux 环境

安装包

对于大多数 Linux 发行版本来说，可以使用官方推荐的方法：

```
Debian/Ubuntu
apt-get install memcached
```

```
Redhat/Fedora/CentOS
yum install memcached
```

2. Win 环境

1. 下载 memcache 的 windows 稳定版

下载地址 <http://jehiah.cz/projects/memcached-win32/files/memcached-1.2.1-win32.zip>

解压放某个盘下面，比如在 c:\memcached

3. linux 启动 memcached

启动参数说明：

```
-d 选项是启动一个守护进程
-m 是分配给 Memcache 使用的内存数量，单位是 MB，默认 64MB
-M return error on memory exhausted (rather than removing items)
```

-u 是运行 Memcache 的用户，如果当前为 root 的话，需要使用此参数指定用户

-l 是监听的服务器 IP 地址，默认为所有网卡

-p 是设置 Memcache 的 TCP 监听的端口，最好是 1024 以上的端口

-c 选项是最大运行的并发连接数，默认是 1024

-P 是设置保存 Memcache 的 pid 文件

-f chunk size growth factor (default: 1.25)

-I Override the size of each slab page. Adjusts max item size(1.4.2 版本新增)

```
/usr/local/memcached/bin/memcached -d -m 100 -c 1000 -u root -p 11211
```

可以启动多个守护进程，但是端口不能重复。设置开机启动的话可以将上行命令增加到 /etc/rc.d/rc.local 文件中。

4. windows 下启动

在终端（也即 cmd 命令界面）cd 到解压目录（这里是 c:\memcached），运行 memcached.exe -d install 安装服务

运行 memcached.exe -d start，memcached 会使用默认的端口(11211)来启动，你可以在任务管理器中看到 memcached.exe

-p 监听的端口

-l 连接的 IP 地址， 默认是本机

-d start 启动 memcached 服务

-d restart 重起 memcached 服务

-d stop|shutdown 关闭正在运行的 memcached 服务

-d install 安装 memcached 服务

-d uninstall 卸载 memcached 服务

-u 以的身份运行（仅在以 root 运行的时候有效）

-m 最大内存使用，单位 MB。默认 64MB

-M 内存耗尽时返回错误，而不是删除项

-c 最大同时连接数， 默认是 1024

-f 块大小增长因子， 默认是 1.25

-n 最小分配空间，key+value+flags 默认是 48

-h 显示帮助

2) Python 操作 Memcached

memcached API 地址 <http://code.google.com/p/memcached/wiki/Clients>

网上流传说 Python-API 中效率最高的是 python-libmemcached，这里居然看到了 hongqn(豆瓣首席架构师，后来也得到证实 python-libmemcached 是豆瓣贡献)，看来豆瓣的阳光真的是撒满了 Python 的各个角落。

另外还有 python-memcached (100%纯 Python)，python-memcache (据说有内存泄漏问题？)，cmemcache (代码多年未更新？)。

不过由于目前需要中效率并没有太高要求，于是选择了使用最多的 **python-memcached**:

安装 **python-memcached**

```
easy_install python-memcached
```

1. Python 操作 memcached

```
import memcache  
mc = memcache.Client(['127.0.0.1:11211'],debug=True)  
mc.set('name','luo',60)  
print mc.get('name')
```

2. Memcached 常用方法

memcache 其实是一个 **map** 结构，最常用的几个函数：

保存数据

set(key,value,timeout) 把 **key** 映射到 **value**, **timeout** 指的是什么时候这个映射失效

add(key,value,timeout) 仅当存储空间中不存在键相同的数据时才保存

replace(key,value,timeout) 仅当存储空间中存在键相同的数据时才保存

获取数据

get(key) 返回 **key** 所指向的 **value**

get_multi([key1,key2,key3,key4]) 可以非同步地同时取得多个键值，比循环调用 **get** 快数十倍

删除数据

delete(key, timeout) 删除键为 **key** 的数据, **timeout** 为时间值, 禁止在 **timeout** 时间内名为 **key** 的键保存新数据 (**set** 函数无效)。

二十七、 SQLite 模块

1) 简单的介绍

SQLite 数据库是一款非常小巧的嵌入式开源数据库软件，也就是说没有独立的维护进程，所有的维护都来自于程序本身。它是遵守 **ACID** 的关联式数据库管理系统，它的设计目标是嵌入式的，而且目前已经有很多嵌入式产品中使用了它，它占用资源非常的低，在嵌入式设备中，可能只需要几百 K 的内存就够了。它能够支持 **Windows/Linux/Unix** 等等主流的操作系统，同时能够跟很多程序语言相结合，比如 **Tcl**、**C#**、**PHP**、**Java** 等，还有 **ODBC** 接口，同样比起 **Mysql**、**PostgreSQL** 这两款开源世界著名的数据库管理系统来讲，它的处理速度比他们都快。**SQLite** 第一个 **Alpha** 版本诞生于 2000 年 5 月。至今已经有 10 个年头，**SQLite** 也迎来了一个版本 **SQLite 3** 已经发布。

2) 安装与使用

1. 导入 Python SQLite 数据库模块

Python2.5 之后，内置了 `SQLite3`，成为了内置模块，这给我们省了安装的功夫，只需导入即可~

```
import sqlite3
```

2. 创建/打开数据库

在调用 `connect` 函数的时候，指定库名称，如果指定的数据库存在就直接打开这个数据库，如果不存在就新创建一个再打开。

```
cx = sqlite3.connect("E:/test.db")
```

也可以创建数据库在内存中。

```
con = sqlite3.connect(":memory:")
```

3. 数据库连接对象

打开数据库时返回的对象 `cx` 就是一个数据库连接对象，它可以有以下操作：

- `commit()`--事务提交
- `rollback()`--事务回滚
- `close()`--关闭一个数据库连接
- `cursor()`--创建一个游标

关于 `commit()`，如果 `isolation_level` 隔离级别默认，那么每次对数据库的操作，都需要使用该命令，你也可以设置 `isolation_level=None`，这样就变为自动提交模式。

4. 使用游标查询数据库

我们需要使用游标对象 SQL 语句查询数据库，获得查询对象。通过以下方法来定义一个游标。

```
cu=cx.cursor()
```

- 游标对象有以下的操作：
- `execute()`--执行 sql 语句
- `executemany()`--执行多条 sql 语句
- `close()`--关闭游标
- `fetchone()`--从结果中取一条记录，并将游标指向下一条记录
- `fetchmany()`--从结果中取多条记录
- `fetchall()`--从结果中取出所有记录
- `scroll()`--游标滚动

1. 建表

```
cu.execute("create table catalog (id integer primary key,pid integer,name varchar(10) UNIQUE,nickname text NULL)")
```

上面语句创建了一个叫 `catalog` 的表，它有一个主键 `id`，一个 `pid`，和一个 `name`，`name` 是不可以重复的，以及一个 `nickname` 默认为 `NULL`。

2. 插入数据

请注意避免以下写法：

```
# Never do this -- insecure 会导致注入攻击
pid=200
c.execute("... where pid = '%s'" % pid)
```

正确的做法如下，如果 `t` 只是单个数值，也要采用 `t=(n,)` 的形式，因为元组是不可变的。

```
for t in[(0,10,'abc','Yu'),(1,20,'cba','Xu')]:
```

```
cx.execute("insert into catalog values (?,?,?,?,?)", t)
```

简单的插入两行数据，不过需要提醒的是，只有提交了之后，才能生效。我们使用数据库连接对象 `cx` 来进行提交 `commit` 和回滚 `rollback` 操作。

```
cx.commit()
```

3. 查询

```
cu.execute("select * from catalog")
```

要提取查询到的数据，使用游标的 `fetch` 函数，如：

```
In [10]: cu.fetchall()
Out[10]: [(0, 10, u'abc', u'Yu'), (1, 20, u'cba', u'Xu')]
```

如果我们使用 `cu.fetchone()`，则首先返回列表中的第一项，再次使用，则返回第二项，依次下去。

4. 修改

```
In [12]: cu.execute("update catalog set name='Boy' where id = 0")
In [13]: cx.commit()
```

注意，修改数据以后提交

5. 删除

```
cu.execute("delete from catalog where id = 1")
cx.commit()
```

6. 使用中文

请先确定你的 IDE 或者系统默认编码是 `utf-8`，并且在中文前加上 `u`

```
x=u'鱼'
cu.execute("update catalog set name=? where id = 0",x)
cu.execute("select * from catalog")
cu.fetchall()
[(0, 10, u'\u9c7c', u'Yu'), (1, 20, u'cba', u'Xu')]
```

如果要显示出中文字体，那需要依次打印出每个字符串

```
In [26]: for item in cu.fetchall():
....:     for element in item:
....:         print element,
....:     print
....:
0 10 鱼 Yu
1 20 cba Xu
```

游标提供了一种对从表中检索出的数据进行操作的灵活手段，就本质而言，游标实际上是一种能从包括多条数据记录的结果集中每次提取一条记录的机制。游标总是与一条 SQL 选择语句相关联。因为游标由结果集（可以是零条、一条或由相关的选择语句检索出的多条记录）和结果集中指向特定记录的游标位置组成。当决定对结果集进行处理时，必须声明一个指向该结果集的游标。如果曾经用 C 语言写过对文件进行处理的程序，那么游标就像您打开文件所得到的文件句柄一样，只要文件打开成功，该文件句柄就可代表该文件。对于游标而言，其道理是相同的。可见游标能够实现按与传统程序读取平面文件类似的方式处理来自基础表的结果集，从而把表中数据以平面文件的形式呈现给程序。

我们知道关系数据库管理系统实质是面向集合的，在 `Sqlite` 中并没有一种描述表中单一记录的表达形式，除非使用 `where` 子句来限制只有一条记录被选中。因此我们必须借助于游标来进行面向单条记录的数据处理。由此可见，游标允许应用程序对查询语句 `select` 返回的行结果集中每一行进行相同或不同的操作，而不是一次对整个结果集进行同一种操作；它还提供对基于游标位置而对表中数据进行删除或更新的能力；正是游标把作为面向集合的数据库管理系统和面向行的程序设计两者联系起来，使两个数据处理方式能够进行沟通。

3) 一个简单的例子

```
#coding:utf-8
import sqlite3
import sys

def sqlite_basic():
    # Connect to db
    conn = sqlite3.connect('test1.db')
    # create cursor
    c = conn.cursor()
    # Create table
    c.execute('''
        create table if not exists stocks
        (date text, trans text, symbol text,
        qty real, price real)
        ''')
    # Insert a row of data
    c.execute('''
        insert into stocks
        values ('2006-01-05','BUY','楚',100,35.14)
        ''')
    # query the table
    rows = c.execute("select * from stocks")
    # print the table
    for row in rows:
        print(row)
        # delete the row
    c.execute("delete from stocks where symbol=='REHT'")
    # Save (commit) the changes
    conn.commit()
    # Close the connection
    conn.close()

def sqlite_adv():
    conn = sqlite3.connect('test2.db')
    c = conn.cursor()
    c.execute('''
        create table if not exists employee
        (id text, name text, age integer)
        ''')
    # insert many rows
    for t in [('1', 'itech', 10),
              ('2', 'jason', 10),
```

```
('3', 'jack', 30),
]:
    c.execute('insert into employee values (?,?,?)', t)
# create index
create_index = 'CREATE INDEX IF NOT EXISTS idx_id ON employee (id);'
c.execute(create_index)
# more secure
t = ('jason',)
c.execute('select * from employee where name=?', t)
# fetch query result
for row in c.fetchall():
    print(row)
conn.commit()
conn.close()

def sqlite_adv2():
    reload(sys)
    sys.setdefaultencoding('utf-8')
    # memory db
    con = sqlite3.connect(":memory:")
    cur = con.cursor()
    # execute sql
    cur.executescript('''
create table book(
    title,
    author,
    published
);
insert into book(title, author, published)
values (
    'AAA book',
    'Douglas Adams', "kkk");
''')
    rows = cur.execute("select * from book")
    for row in rows:
        print("title:" + row[0])
        print("author:" + row[1])
        print("published:" + str(row[2]))


def sqlite_adv3():
    import datetime

    # Converting SQLite values to custom Python types
    # Default adapters and converters for datetime and timestamp
    con = sqlite3.connect(":memory:", detect_types=sqlite3.PARSE_DECLTYPES |
    sqlite3.PARSE_COLNAMES)
```

```
cur = con.cursor()
cur.execute("create table test(d date, ts timestamp)")

today = datetime.date.today()
now = datetime.datetime.now()

cur.execute("insert into test(d, ts) values (?, ?)", (today, now))
cur.execute("select d, ts from test")
row = cur.fetchone()
print today, "=>", row[0], type(row[0])
print now, "=>", row[1], type(row[1])

cur.execute('select current_date as "d [date]", current_timestamp as "ts [timestamp]"'
from test')
row = cur.fetchone()
print "current_date", row[0], type(row[0])
print "current_timestamp", row[1], type(row[1])

sqlite_basic()
```

4) 中文处理

二十八、 正则表达式

1) 简介

编写验证规则最流行和最简单的方法就是正则表达了，但唯一的一个问题是正则表达式的语法太隐晦了，让人蛋疼无比。很多开发者为了在项目中应用复杂的验证，经常要使用一些小抄来记住正则式的复杂语法和各种常用命令。



也许你是初学者，那以防万一，我先来讲讲什么是正则表达式吧：

正则表达式可以帮助我们更好的描述复杂的文本格式。一旦你描述清楚了这些格式，那你就可以利用它们对文本数据进行检索、替换、提取和修改操作。

下面有一个正则表达式的简单例子。第一步先要引入有关正则式的包：

```
# -*- coding: utf-8 -*-
import re
```

第二步就是用指定的正则式构建一个正则表达式对象，下面的正则式是用来搜索长度为 10 的 a-z 的英文字母：

```
# -*- coding: utf-8 -*-
import re
regex = re.compile('[a-z]{10}')
```

最后，根据正则式在指定数据中检索匹配项，如果匹配 `IsMatch` 方法就会返回 `true`。

```
# -*- coding: utf-8 -*-
import re
regex = re.compile('[a-z]{10}')
m1 = re.match(regex, '1234567890')
m2 = re.match(regex, 'abcdefghijkl')
# m = regex.search("test1234.567").group()
if m1 is not None:
    print 'True'
if m2 is not None:
    print 'True'
```

2) 3 个重要的正则式命令

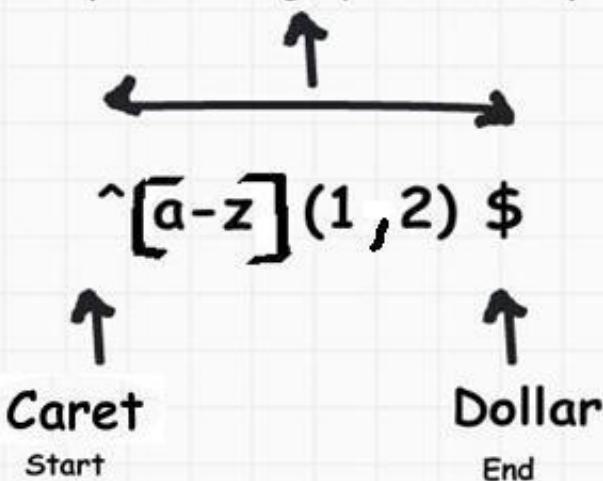
记住正则语法最好的办法就是记住这三样东西：Bracket（括号），caret（插入符号）和 Dollars（美元符号）。



B (括号)	在正则表达式中有 3 种类型的括号 方括号 "[" 和花括号 "{}"。 方括号 "[" 内是需要匹配的字符，花括号 "{}" 内是指定匹配字符的数量。 圆括号 "(" 则是用来分组的。
C (插入符号)	插入符号 "^" 表示正则式的开始。
D (美元符号)	美元符号 "\$" 表示正则式的结束。

现在你知道上面的 3 个语法后，你就可以写世界上任何一条验证规则了。比如下面的例子就很好的说明了上面 3 条正则语法是如何协调运作的。

Validation pattern using square and curly brackets



(注：上图有个错误，"()"应为"{}")

- 上面的这条正则式只能匹配 a-z 的英文字母，同样是在中括号中标明匹配范围。
- 花括号中则是标明匹配字符串的最小长度和最大长度。

- 最后为了让表达式更规则，分别在开头和结尾加上了插入符号"^"和美元符号"\$"。

1. 验证字符串

好了，现在我们就用上面的 3 条语法来实现一些正则表达式的验证规则吧。

检查用户是否输入了 shivkoirala？

```
shivkoirala
```

让我们开始第一个验证，输入的字符在 a-g 之间？

```
[a-g]
```

输入的字符在 a-g 之间并且长度为 3？

```
[a-g]{3}
```

输入的字符在 a-g 之间并且最大长度为 3 最小长度为 1？

```
[a-g]{1,3}
```

我如何在匹配像 91230456, 01237648 那样的固定 8 位数？

```
^[0-9]{8}$
```

如何验证最小长度为 3 最大长度为 7 的数字，如：123, 1274667, 87654？

```
^[0-9]{3,7}$
```

如何验证像 LJI10201020 那样的发票编号，前 3 个是字母剩余为 8 位长度的数字？前三个是字母：

```
^[a-z]{3}
```

后面是 8 位长度的数字：

```
[0-9]{8}
```

所以整个表达式为：

```
^[a-z]{3}[0-9]{8}$
```

验证像 INV19020303 或 inv82083003 那样的前 3 位是不区分大小写的英文字母，剩余 8 位是数字在前面的表达式中只能匹配前 3 个是小写英文字母的发票编号，如果我们输入大写字母那就不能匹配了。所以为了确保前 3 个字母是不区分大小写的，我们就要用表达式^[a-zA-Z]{3}。

完整的正则式如下：

```
^[a-zA-Z]{3}[0-9]{8}$
```

2. 验证 url

我们可以验证简单的网址 URL 格式吗？

第一步：检查是否存在 www：

```
^www.
```

第二步：域名必须是长度在 1-15 的英文字母：

```
.[a-z]{1,15}
```

第三部：以 .com 或者 .org 结束：

```
.(com|org)$
```

完整的表达式如下：

```
^www.[.][a-z]{1,15}[.](com|org)$
```

3. 验证 Email

让我们来看看 BCD（其实也就是上面说的 3 条基本语法）如何验证 email 格式

第一步：email 开头是长度在 1-10 的英文字母，最后跟一个 "@"：

```
^[a-zA-Z]{1,10}@\n
```

第二步：@后面是长度在 1-10 的英文字母，后面跟一个 "."：

```
[a-zA-Z]{1,10}.\n
```

第三步：最后以 .com 或 .org 结束：

```
.(com|org)$\n
```

最后完整的表达式如下：

```
^[a-zA-Z]{1,10}@[a-zA-Z]{1,10}.(com|org)$\n
```

4. 验证值在 0-25 的数字

```
^(([0-9])|([0-1][0-9])|([0-2][0-5]))$\n
```

5. 验证格式为 MM/DD/YYYY, YYYY/MM/DD and DD/MM/YYYY 的日期

步骤	正则式	描述说明
先来检查 DD. 首先 DD 的长度为 1-29 (2 月份), 1-30 (月小), 1-31 (月大). 所以 DD 就是 1-9 或 01-09	[1-9] 0[1-9]	允许用户输入 1-9 或者 01-09.
再为 DD 添加匹配 10-19	[1-9] 1[0-9]	允许用户输入 01-19.
再为 DD 添加匹配 20-29	[1-9] 1[0-9] 2[0-9]	允许用户输入 01-29.
i 再为 DD 添加匹配 30-31	[1-9] 1[0-9] 2[0-9] 3[0-1]	最后用户可以输入 01-31.
再来匹配日期间的分隔符 "/", "-"	[/ . -]	允许用户输入日期分隔符.
MM 也是类似的操作	[1-9] 0[1-9] 1[0-2]	让用户输入月份值 01-12.
最后就是 YY 的操作	1[9][0-9][0-9] 2[0][0-9][0-9]	允许用户输入年份 1900-2099.

最后 DD/MM/YYYY 格式的日期的正则表达式为：

```
^(1[9]|0[1-9]|1[0-9]|2[0-9]|3[0-1])[- / .](1[9]|0[1-9]|1[0-2])[- / .](1[9][0-9][0-9]|2[0][0-9][0-9])$\n
```

MM/DD/YYYY 格式的日期:

```
^(1[0-9]|2[0-9]|3[0-9])[0-9]{2}(/)(1[0-9]|2[0-9]|3[0-9])[0-9]{2}(/)(1[0-9]|2[0-9]|3[0-9])[0-9]{2}$
```

YYYY/MM/DD 格式的日期:

```
^(1[0-9]{4})(/)(1[0-9]|2[0-9]|3[0-9])[0-9]{2}(/)(1[0-9]|2[0-9]|3[0-9])[0-9]{2}(/)(1[0-9]|2[0-9]|3[0-9])[0-9]{2}$
```

3) 正则表达符

1. 你常用的元字符

代码	说明
.	匹配除换行符以外的任意字符
\w	匹配字母或数字或下划线或汉字
\s	匹配任意的空白符
\d	匹配数字
\b	匹配单词的开始或结束
^	匹配字符串的开始
\$	匹配字符串的结束

比如一个网站如果要求你填写的 QQ 号必须为 5 位到 12 位数字时，可以使用:

```
^\d{5,12}$
```

2. 常用的限定符

代码	说明
*	重复零次或更多次
+	重复一次或更多次
?	重复零次或一次
{n}	重复 n 次
{n,}	重复 n 次或更多次
{n,m}	重复 n 到 m 次

下面是一些使用重复的例子:

Windows\d+ 匹配 Windows 后面跟 1 个或更多数字

^\w+ 匹配一行的第一个单词(或整个字符串的第一个单词，具体匹配哪个意思得看选项设置)

```
^:匹配字符串的开始  
$:匹配字符串的结尾  
\b:匹配一个单词的边界  
\d:匹配任意数字  
\D:匹配任意非数字字符  
x?:匹配一个可选的 x 字符 (换句话说，它匹配 1 次或者 0 次 x 字符)
```

x*: 匹配 0 次或者多次 x 字符

x+: 匹配 1 次或者多次 x 字符

x{n,m}: 匹配 x 字符，至少 n 次，至多 m 次

{a|b|c}: 要么匹配 a，要么匹配 b，要么匹配 c

(x): 一般情况下表示一个记忆组(remembered group)

你可以利用 re.search 函数返回对象的 groups() 函数获取它的值

3. 字符串类

要想查找数字，字母或数字，空白是很简单的，因为已经有了对应这些字符集合的元字符，但是如果你想匹配没有预定义元字符的字符集合(比如元音字母 a,e,i,o,u)，应该怎么办？

很简单，你只需要在方括号里列出它们就行了，像 [aeiou] 就匹配任何一个英文元音字母，[.?!] 匹配标点符号(.或?或!)。

我们也可以轻松地指定一个字符范围，像 [0-9] 代表的含意与 \d 就是完全一致的：一位数字；同理 [a-zA-Z] 也完全等同于 \w (如果只考虑英文的话)。

如果我们要匹配一个电话区号：像 (010)88886666，或 022-22334455，或 02912345678 等

第一步：前面的区号可能有括号也可能没有，区号与电话号码之间可能有 - 号也可能没有

首先是一个转义字符 \(), 它能出现 0 次或 1 次 ?

然后是一个 0，后面跟着 2 个数字 \d{2}

然后是) 或 - 或空格中的一个，它出现 1 次或不出现 ?，最后是 8 个数字 \d{8}

\(?0\d{2}[]-]

第二步：匹配后面的电话

\(?0\d{2}[]-]? \d{8}

“(” 和 “)” 也是元字符，后面的分组节里会提到，所以在这里需要使用转义。

这个表达式可以匹配几种格式的电话号码，像 (010)88886666，或 022-22334455，或 02912345678 等。我们对它进行一些分析吧：首先是一个转义字符 \(), 它能出现 0 次或 1 次 (?), 然后是一个 0, 后面跟着 2 个数字 \d{2}, 然后是) 或 - 或空格中的一个，它出现 1 次或不出现 (?), 最后是 8 个数字 (\d{8})。

```
# -*- coding: utf-8 -*-
import re

regex = re.compile('\(?0\d{2}[]-]? \d{8}')
m1 = re.search(regex, '(010)88886666')
m2 = re.search(regex, '022-22334455')
m3 = re.search(regex, '02912345678')
m4 = re.search(regex, '(010)88886666 022-22334455 02912345678')
if m4 is not None:
    print m4.group()
m5 = re.findall(regex, '(010)88886666 022-22334455 02912345678')
print m5
```

4. 分枝条件

不幸的是，刚才那个表达式也能匹配 **010)12345678** 或 **(022-87654321** 这样的“不正确”的格式。

要解决这个问题，我们需要用到分枝条件。正则表达式里的分枝条件指的是有几种规则，如果满足其中任意一种规则都应该当成匹配，具体方法是用 | 把不同的规则分隔开。听不明白？没关系，看例子：

0\d{2}-\d{8} | 0\d{3}-\d{7} 这个表达式能匹配两种以连字号分隔的电话号码：一种是三位区号，

8位本地号如 **010-12345678**，一种是4位区号，7位本地号 **0376-2233445**

\((0\d{2})\)[-]?\d{8} | 0\d{2}[-]?\d{8} 这个表达式匹配3位区号的电话号码，其中

区号可以用小括号括起来，也可以不用，区号与本地号间可以用连字号或空格间隔，也可以没有间隔。你可以试试用分枝条件把这个表达式扩展成也支持4位区号的。

\d{5}-\d{4} | \d{5} 这个表达式用于匹配美国的邮政编码。美国邮编的规则是 5 位数字，或者用连字

号间隔的 9 位数字。之所以要给出这个例子是因为它能说明一个问题：使用分枝条件时，要注意各个条件的顺序。

如果你把它改成 **\d{5} | \d{5}-\d{4}** 的话，那么就只会匹配 5 位的邮编(以及 9 位邮编的前 5 位)。原因是匹配分枝条件时，将会从左到右地测试每个条件，如果满足了某个分枝的话，就不会去再管其它的条件了。

5. 分组

我们已经提到了怎么重复单个字符（直接在字符后面加上限定符就行了）；但如果想要重复多个字符又该怎么办？你可以用小括号来指定子表达式(也叫做分组)，然后你就可以指定这个子表达式的重复次数了，你也可以对子表达式进行其它一些操作(后面会有介绍)。

(\d{1,3}\.){3}\d{1,3} 是一个简单的 IP 地址匹配表达式。要理解这个表达式，请按下列

顺序分析它：

\d{1,3} 匹配 1 到 3 位的数字

(\d{1,3}\.){3} 匹配三位数字加上一个英文句号(这个整体也就是这个分组)重复 3 次

最后再加上一个一到三位的数字 **(\d{1,3})**

IP 地址中每个数字都不能大于 255

不幸的是，它也将匹配 **256.300.888.999** 这种不可能存在的 IP 地址。如果能使用算术比较的话，或许能简单地解决这个问题，但是正则表达式中并不提供关于数学的任何功能，所以只能使用冗长的分组，选择，字符类来描述一个正确的 IP 地址：

((2[0-4]\d|25[0-5]| [01]?\d\d?)\.\){3}(2[0-4]\d|25[0-5]| [01]?\d\d?)

理解这个表达式的关键是理解 **`2[0-4]\d|25[0-5]|01?\d\d?`**, 这里我就不细说了，你自己应该能分析得出来它的意义。

6. 反义

有时需要查找不属于某个能简单定义的字符类的字符。比如想查找除了数字以外，其它任意字符都行的情况，这时需要用到反义：

表 3. 常用的反义代码	
代码/语法	说明
<code>\W</code>	匹配任意不是字母，数字，下划线，汉字的字符
<code>\S</code>	匹配任意不是空白符的字符
<code>\D</code>	匹配任意非数字的字符
<code>\B</code>	匹配不是单词开头或结束的位置
<code>[^x]</code>	匹配除了 x 以外的任意字符
<code>[^aeiou]</code>	匹配除了 aeiou 这几个字母以外的任意字符

例子：

`\S+` 匹配不包含空白符的字符串

`<a[^>]+>` 匹配用尖括号括起来的以 a 开头的字符串

7. 后向引用

使用小括号指定一个子表达式后，**匹配这个子表达式的文本**(也就是此分组捕获的内容)可以在表达式或其它程序中作进一步的处理。默认情况下，每个分组会自动拥有一个**组号**，规则是：从左向右，以分组的左括号为标志，第一个出现的分组的组号为 1，第二个为 2，以此类推。

- 分组 0 对应整个正则表达式
- 实际上组号分配过程是要从左向右扫描两遍的：第一遍只给未命名组分配，第二遍只给命名组分配——因此所有命名组的组号都大于未命名的组号
- 你可以使用`(?:exp)`这样的语法来剥夺一个分组对组号分配的参与权。

后向引用用于重复搜索前面某个分组匹配的文本。例如，`\1` 代表分组 1 匹配的文本。难以理解？请看示例：
`\b(\w+)\b\s+\1\b` 可以用来匹配**重复的单词**，像 *go go*，或者 *kitty kitty*。这个表达式首先是一个**单词**，也就是**单词开始处和结束处之间的多于一个的字母或数字**(`\b(\w+)\b`)，这个单词会被捕获到编号为 1 的分组中，然后是**1个或几个空白符**(`\s+`)，最后是分组 1 中捕获的内容（也就是前面匹配的那个单词）(`\1`)。

你也可以自己指定子表达式的**组名**。要指定一个子表达式的组名，请使用这样的语法：`(?<Word>\w+)`(或者把尖括号换成'也行：`(?'Word'\w+)`)，这样就把`\w+`的组名指定为 **Word** 了。要反向引用这个分组捕获的内容，你可以使用`\k<Word>`，所以上一个例子也可以写成这样：`\b(?<Word>\w+)\b\s+\k<Word>\b`。

使用小括号的时候，还有很多特定用途的语法。下面列出了最常用的一些：

表 4. 常用分组语法

分类	代码/语法	说明
捕获	<code>(exp)</code>	匹配 exp，并捕获文本到自动命名的组里
	<code>(?<name>exp)</code>	匹配 exp，并捕获文本到名称为 name 的组里，也可以写成 <code>(?'name'exp)</code>
	<code>(?:exp)</code>	匹配 exp，不捕获匹配的文本，也不给此分组分配组号
零宽断言	<code>(?=exp)</code>	匹配 exp 前面的位置

	<code>(?<=exp)</code>	匹配 <code>exp</code> 后面的位置
	<code>(?!exp)</code>	匹配后面跟的不是 <code>exp</code> 的位置
	<code>(?<!exp)</code>	匹配前面不是 <code>exp</code> 的位置
注释	<code>(?#comment)</code>	这种类型的分组不对正则表达式的处理产生任何影响，用于提供注释让人阅读

我们已经讨论了前两种语法。第三个(`?>exp`)不会改变正则表达式的处理方式，只是这样的组匹配的内容不会像前两种那样被捕获到某个组里面，也不会拥有组号。“我为什么会想要这样做？”—好问题，你觉得为什么呢？

8. 贪婪与懒惰

当正则表达式中包含能接受重复的限定符时，通常的行为是（在使整个表达式能得到匹配的前提下）匹配尽可能多的字符。以这个表达式为例：`a.*b`，它将会匹配最长的以 a 开始，以 b 结束的字符串。如果用它来搜索 `aabab` 的话，它会匹配整个字符串 `aabab`。这被称为**贪婪匹配**。

有时，我们更需要**懒惰匹配**，也就是匹配尽可能少的字符。前面给出的限定符都可以被转化为懒惰匹配模式，只要在它后面加上一个问号`?`。这样`.*?`就意味着匹配任意数量的重复，但是在能使整个匹配成功的前提下使用最少的重复。现在看看懒惰版的例子吧：

`a.*?b` 匹配最短的，以 a 开始，以 b 结束的字符串。如果把它应用于 `aabab` 的话，它会匹配 `aab`（第一到第三个字符）和 `ab`（第四到第五个字符）。

为什么第一个匹配是 `aab`（第一到第三个字符）而不是 `ab`（第二到第三个字符）？简单地说，因为正则表达式有另一条规则，比懒惰 / 贪婪规则的优先级更高：最先开始的匹配拥有最高的优先权—**The match that begins earliest wins**。

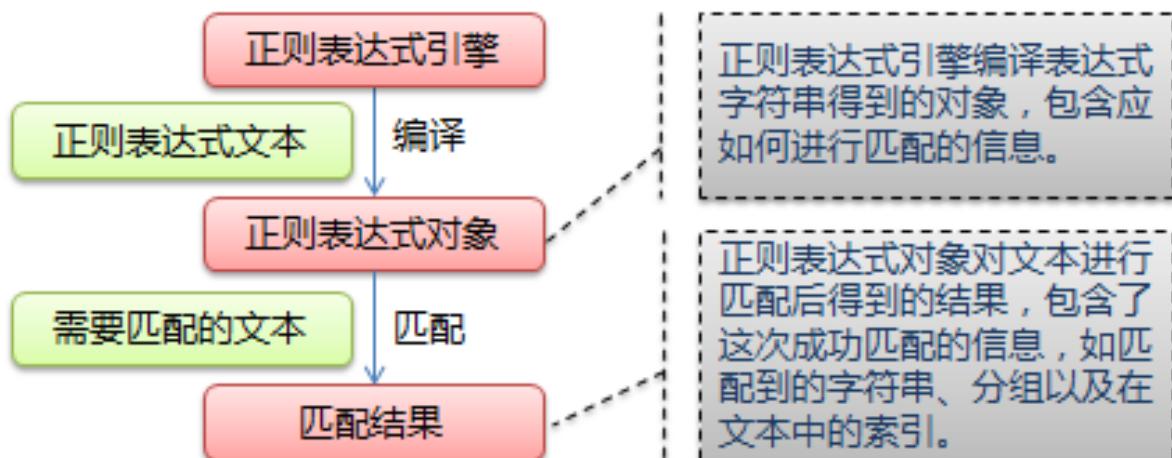
表 5. 懒惰限定符

代码/语法	说明
<code>*?</code>	重复任意次，但尽可能少重复
<code>+?</code>	重复 1 次或更多次，但尽可能少重复
<code>??</code>	重复 0 次或 1 次，但尽可能少重复
<code>{n,m}?</code>	重复 n 到 m 次，但尽可能少重复
<code>{n,}?</code>	重复 n 次以上，但尽可能少重复

4) 在 Python 中使用

正则表达式并不是 `Python` 的一部分。正则表达式是用于处理字符串的强大工具，拥有自己独特的语法以及一个独立的处理引擎，效率上可能不如 `str` 自带的方法，但功能十分强大。得益于这一点，在提供了正则表达式的语言里，正则表达式的语法都是一样的，区别只在于不同的编程语言实现支持的语法数量不同；但不用担心，不被支持的语法通常是不常用的部分。如果已经在其他语言里使用过正则表达式，只需要简单看一看就可以上手了。

下图展示了使用正则表达式进行匹配的流程：



正则表达式的大致匹配过程是：依次拿出表达式和文本中的字符比较，如果每一个字符都能匹配，则匹配成功；一旦有匹配不成功的字符则匹配失败。如果表达式中有量词或边界，这个过程会稍微有一些不同，但也是很好理解的，看下图中的示例以及自己多使用几次就能明白。

下图列出了 Python 支持的正则表达式元字符和语法：

语法	说明	表达式实例	完整匹配的字符串
字符			
一般字符	匹配自身	abc	abc
.	匹配任意除换行符"\n"外的字符。 在DOTALL模式中也能匹配换行符。	a.c	abc
\	转义字符，使后一个字符改变原来的意思。 如果字符串中有字符*需要匹配，可以使用*或者字符集[*]。	a\c a\\c	a.c a\c
[...]	字符集（字符类）。对应的位置可以是字符集中任意字符。 字符集中的字符可以逐个列出，也可以给出范围，如[abc]或[a-c]。第一个字符如果是^则表示取反，如[^abc]表示不是abc的其他字符。 所有的特殊字符在字符集中都失去其原有的特殊含义。在字符集中如果要使用]、-或^，可以在前面加上反斜杠，或把]、-放在第一个字符，把^放在非第一个字符。	a[bcd]e	abe ace ade
预定义字符集（可以在字符集 [...] 中）			
\d	数字：[0-9]	a\dc	a1c
\D	非数字：[^0-9]	a\Dc	abc
\s	空白字符：[<空格>\t\r\n\f\v]	a\s c	a c
\S	非空白字符：[^0-9]	a\S c	abc
\w	单词字符：[A-Za-z0-9_]	a\wc	abc
\W	非单词字符：[^0-9_]	a\Wc	a c
数量词（用在字符或(...)之后）			
*	匹配前一个字符0或无限次。	abc*	ab abccc
+	匹配前一个字符1次或无限次。	abc+	abc abccc
?	匹配前一个字符0次或1次。	abc?	ab abc
{m}	匹配前一个字符m次。	ab{2}c	abbc
{m,n}	匹配前一个字符m至n次。 m和n可以省略：若省略m，则匹配0至n次；若省略n，则匹配m至无限次。	ab{1,2}c	abc abbc
*? +? ??	使 * + ? {m,n} 变成非贪婪模式。		示例将在下文中介绍。
{m,n}?			

边界匹配 (不消耗待匹配字符串中的字符)			
^	匹配字符串开头。 在多行模式中匹配每一行的开头。	^abc	abc
\$	匹配字符串末尾。 在多行模式中匹配每一行的末尾。	abc\$	abc
\A	仅匹配字符串开头。	\Aabc	abc
\Z	仅匹配字符串末尾。	abc\Z	abc
\b	匹配\w和\W之间。	a\b!bc	a!bc
\B	[^\b]	a\Bbc	abc
逻辑、分组			
	代表左右表达式任意匹配一个。 它总是先尝试匹配左边的表达式，一旦成功匹配则跳过匹配右边的表达式。 如果 没有被包括在()中，则它的范围是整个正则表达式。	abc def	abc def
(...)	被括起来的表达式将作为分组，从表达式左边开始每遇到一个分组的左括号'('，编号+1。 另外，分组表达式作为一个整体，可以后接数量词。表达式中的 仅在该组中有效。	(abc){2}	abcabc a456c
(?P<name> ...)	分组，除了原有的编号外再指定一个额外的别名。	(?P<id>abc){2}	abcabc
\<number>	引用编号为<number>的分组匹配到的字符串。	(\d)abc\1	1abc1 5abc5
(?P=name)	引用别名为<name>的分组匹配到的字符串。	(?P<id>\d)abc(?P=id)	1abc1 5abc5
特殊构造 (不作为分组)			
(?:...)	(...)的不分组版本，用于使用' '或后接数量词。	(?:abc){2}	abcabc
(?iLmsux)	iLmsux的每个字符代表一个匹配模式，只能用在正则表达式的开头，可选多个。匹配模式将在下文中介绍。	(?i)abc	AbC
(#...)	#后的內容将作为注释被忽略。	abc(#comment)123	abc123
(?=...)	之后的字符串内容需要匹配表达式才能成功匹配。 不消耗字符串内容。	a(?=\d)	后面是数字的a
(?!...)	之后的字符串内容需要不匹配表达式才能成功匹配。 不消耗字符串内容。	a(?! \d)	后面不是数字的a
(?<=...)	之前的字符串内容需要匹配表达式才能成功匹配。 不消耗字符串内容。	(?<=\d)a	前面是数字的a
(?<!...)	之前的字符串内容需要不匹配表达式才能成功匹配。 不消耗字符串内容。	(?<!\d)a	前面不是数字的a
(?i(id/name) yes-pattern no-pattern)	如果编号为id/别名为name的组匹配到字符，则需要匹配yes-pattern，否则需要匹配no-pattern。 no-patern可以省略。	(\d)abc(?:1)\d abc)	1abc2 abcabc http://www.cnblogs.com/huxi

5) Re 模块

Python 通过 `re` 模块提供对正则表达式的支持。使用 `re` 的一般步骤是先将正则表达式的字符串形式编译为 `Pattern` 实例，然后使用 `Pattern` 实例处理文本并获得匹配结果（一个 `Match` 实例），最后使用 `Match` 实例获得信息，进行其他的操作。

```
# encoding: UTF-8
import re

# 将正则表达式编译成 Pattern 对象
pattern = re.compile(r'hello')

# 使用 Pattern 匹配文本，获得匹配结果，无法匹配时将返回 None
match = pattern.match('hello world!')

if match:
    # 使用 Match 获得分组信息
    print match.group()

### 输出 ###
# hello
```

re.compile(strPattern[, flag]):

这个方法是 **Pattern** 类的工厂方法，用于将字符串形式的正则表达式编译为 **Pattern** 对象。第二个参数 **flag** 是匹配模式，取值可以使用按位或运算符'|'表示同时生效，比如 **re.I | re.M**。另外，你也可以在 **regex** 字符串中指定模式，比如 **re.compile('pattern', re.I | re.M)** 与 **re.compile('(?im)pattern')** 是等价的。可选值有：

- **re.I(re.IGNORECASE)**: 忽略大小写（括号内是完整写法，下同）
- **M(MULTILINE)**: 多行模式，改变'^'和'\$'的行为（参见上图）
- **S(DOTALL)**: 点任意匹配模式，改变'.'的行为
- **L(LOCALE)**: 使预定字符类 \w \W \b \B \s \S 取决于当前区域设定
- **U(UNICODE)**: 使预定字符类 \w \W \b \B \s \S \d \D 取决于 **unicode** 定义的字符属性
- **X(VERBOSE)**: 详细模式。这个模式下正则表达式可以是多行，忽略空白字符，并可以加入注释。以下两个正则表达式是等价的：

```
# encoding: UTF-8
import re

a = re.compile(r"""\d + # the integral part
                    \. # the decimal point
                    \d * # some fractional digits""", re.X)
b = re.compile(r"\d+\.\d*")
```

re 提供了众多模块方法用于完成正则表达式的功能。这些方法可以使用 **Pattern** 实例的相应方法替代，唯一的好处是少写一行 **re.compile()** 代码，但同时也无法复用编译后的 **Pattern** 对象。这些方法将在 **Pattern** 类的实例方法部分一起介绍。如上面这个例子可以简写为：

```
m = re.match(r'hello', 'hello world!')
print m.group()
```

re 模块还提供了一个方法 **escape(string)**，用于将 **string** 中的正则表达式元字符如`*`/`+`/`?`等之前加上转义符再返回，在需要大量匹配元字符时有那么一点用。

6) Match 方法

Match 对象是一次匹配的结果，包含了很多关于此次匹配的信息，可以使用 **Match** 提供的可读属性或方法来获取这些信息。

属性：

- **string**: 匹配时使用的文本。
- **re**: 匹配时使用的 **Pattern** 对象。
- **pos**: 文本中正则表达式开始搜索的索引。值与 **Pattern.match()** 和 **Pattern.search()** 方法的同名参数相同。
- **endpos**: 文本中正则表达式结束搜索的索引。值与 **Pattern.match()** 和 **Pattern.search()** 方法的同名参数相同。
- **lastindex**: 最后一个被捕获的分组在文本中的索引。如果没有被捕获的分组，将为 **None**。
- **lastgroup**: 最后一个被捕获的分组的别名。如果这个分组没有别名或者没有被捕获的分组，将为 **None**。

方法：

- **group([group1, ...])**:

获得一个或多个分组截获的字符串；指定多个参数时将以元组形式返回。**group1** 可以使用编号也可以使用别名；编号 0 代表整个匹配的子串；不填写参数时，返回 **group(0)**；没有截获字符串的组返回 **None**；截获了多次的组返回最后一次截获的子串。

- **groups([default])**:

以元组形式返回全部分组截获的字符串。相当于调用 **group(1,2,...last)**。**default** 表示没有截获字符串的组以这个值替代，默认为 **None**。

- **groupdict([default])**:

返回以有别名的组的别名为键、以该组截获的子串为值的字典，没有别名的组不包含在内。**default** 含义同上。

- **start([group])**:

返回指定的组截获的子串在 **string** 中的起始索引（子串第一个字符的索引）。**group** 默认值为 0。

- **end([group])**:

返回指定的组截获的子串在 **string** 中的结束索引（子串最后一个字符的索引+1）。**group** 默认值为 0。

- **span([group])**:

返回(**start(group)**, **end(group)**)。

- **expand(template)**:

将匹配到的分组代入 **template** 中然后返回。**template** 中可以使用 \id 或 \g<id>、\g<name> 引用分组，但不能使用编号 0。**\id** 与 **\g<id>** 是等价的；但 **\10** 将被认为是第 10 个分组，如果你想表达 \1 之后是字符'0'，只能使用 **\g<1>0**。

```
# encoding: UTF-8
import re
m = re.match(r'(\w+) (\w+)(?P<sign>.* )', 'hello world! sign')

print "m.string:", m.string
print "m.re:", m.re
print "m.pos:", m.pos
print "m.endpos:", m.endpos
print "m.lastindex:", m.lastindex
print "m.lastgroup:", m.lastgroup

print "m.group(1,2):", m.group(1, 2)
print "m.groups():", m.groups()
print "m.groupdict():", m.groupdict()
print "m.start(2):", m.start(2)
print "m.end(2):", m.end(2)
print "m.span(2):", m.span(2)
print r"m.expand(r'\2 \1\3'):", m.expand(r'\2 \1\3')
```

```
### output ###
# m.string: hello world!
# m.re: <_sre.SRE_Pattern object at 0x016E1A38>
# m.pos: 0
# m.endpos: 12
# m.lastindex: 3
# m.lastgroup: sign
# m.group(1,2): ('hello', 'world')
# m.groups(): ('hello', 'world', '!')
# m.groupdict(): {'sign': '!'}
# m.start(2): 6
# m.end(2): 11
# m.span(2): (6, 11)
# m.expand(r'\2 \1\3'): world hello!
```

7) Pattern 方法

`Pattern` 对象是一个编译好的正则表达式，通过 `Pattern` 提供的一系列方法可以对文本进行匹配查找。

`Pattern` 不能直接实例化，必须使用 `re.compile()` 进行构造。

`Pattern` 提供了几个可读属性用于获取表达式的相关信息：

- `pattern`: 编译时用的表达式字符串。
- `flags`: 编译时用的匹配模式。数字形式。
- `groups`: 表达式中分组的数量。
- `groupindex`: 以表达式中有别名的组的别名为键、以该组对应的编号为值的字典，没有别名的组不包含在内。

```
# encoding: UTF-8
import re
p = re.compile(r'(\w+) (\w+)(?P<sign>.*')', re.DOTALL)

print "p.pattern:", p.pattern
print "p.flags:", p.flags
print "p.groups:", p.groups
print "p.groupindex:", p.groupindex

### output ###
# p.pattern: (\w+) (\w+)(?P<sign>.*')
# p.flags: 16
# p.groups: 3
# p.groupindex: {'sign': 3}
```

实例方法[| `re` 模块方法]:

- **`match(string[, pos[, endpos]])` | `re.match(pattern, string[, flags])`:**

这个方法将从 `string` 的 `pos` 下标处起尝试匹配 `pattern`；如果 `pattern` 结束时仍可匹配，则返回一个 `Match` 对象；如果匹配过程中 `pattern` 无法匹配，或者匹配未结束就已到达 `endpos`，则返回 `None`。

pos 和 endpos 的默认值分别为 0 和 len(string); re.match()无法指定这两个参数, 参数 flags 用于编译 pattern 时指定匹配模式。

注意：这个方法并不是完全匹配。当 pattern 结束时若 string 还有剩余字符，仍然视为成功。想要完全匹配，可以在表达式末尾加上边界匹配符'\$'。

- **search(string[, pos[, endpos]]) | re.search(pattern, string[, flags]):**

这个方法用于查找字符串中可以匹配成功的子串。从 string 的 pos 下标处起尝试匹配 pattern, 如果 pattern 结束时仍可匹配，则返回一个 Match 对象；若无法匹配，则将 pos 加 1 后重新尝试匹配；直到 pos=endpos 时仍无法匹配则返回 None。

pos 和 endpos 的默认值分别为 0 和 len(string)); re.search()无法指定这两个参数, 参数 flags 用于编译 pattern 时指定匹配模式。

```
# encoding: UTF-8
import re

# 将正则表达式编译成 Pattern 对象
pattern = re.compile(r'world')

# 使用 search()查找匹配的子串, 不存在能匹配的子串时将返回 None
# 这个例子中使用 match()无法成功匹配
match = pattern.search('hello world!')

if match:
    # 使用 Match 获得分组信息
    print match.group()

### 输出 ###
# world
```

- **split(string[, maxsplit]) | re.split(pattern, string[, maxsplit]):**

按照能够匹配的子串将 string 分割后返回列表。maxsplit 用于指定最大分割次数, 不指定将全部分割。

```
import re

p = re.compile(r'\d+')
print p.split('one1two2three3four4')

### output ###
# ['one', 'two', 'three', 'four', '']
```

- **findall(string[, pos[, endpos]]) | re.findall(pattern, string[, flags]):**

搜索 string, 以列表形式返回全部能匹配的子串。

```
# encoding: UTF-8
import re

p = re.compile(r'\d+')
print p.findall('one1two2three3four4')

### output ###
# ['1', '2', '3', '4']
```

finditer(string[, pos[, endpos]]) | re.finditer(pattern, string[, flags]):

搜索 string，返回一个顺序访问每一个匹配结果（Match 对象）的迭代器。

```
# encoding: UTF-8
import re

p = re.compile(r'\d+')
for m in p.finditer('one1two2three3four4'):
    print m.group(),

### output ####
# 1 2 3 4
```

sub(repl, string[, count]) | re.sub(pattern, repl, string[, count]):

使用 repl 替换 string 中每一个匹配的子串后返回替换后的字符串。

当 repl 是一个字符串时，可以使用\id 或\g<id>、\g<name>引用分组，但不能使用编号 0。

当 repl 是一个方法时，这个方法应当只接受一个参数（Match 对象），并返回一个字符串用于替换（返回的字符串中不能再引用分组）。

count 用于指定最多替换次数，不指定时全部替换。

```
# encoding: UTF-8
import re

p = re.compile(r'(\w+) (\w+)')
s = 'i say, hello world!'

print p.sub(r'\2 \1', s)

def func(m):
    return m.group(1).title() + ' ' + m.group(2).title()

print p.sub(func, s)

### output ####
# say i, world hello!
# I Say, Hello World!
```

subn(repl, string[, count]) | re.subn(pattern, repl, string[, count]):

返回 (sub(repl, string[, count]), 替换次数)。

```
# encoding: UTF-8
import re

p = re.compile(r'(\w+) (\w+)')
s = 'i say, hello world!'

print p.subn(r'\2 \1', s)

def func(m):
    return m.group(1).title() + ' ' + m.group(2).title()
```

```
print p.subn(func, s)
```

```
#### output ####
# ('say i, world hello!', 2)
# ('I Say, Hello World!', 2)
```

```
# -*- coding: utf-8 -*-
import re

#1. 匹配多个字符串
bt = 'bat|bet|bit'
m = re.search(bt, 'He bit bet')
if m is not None:
    print m.group()
m = re.findall(bt, 'He bit bet')
print m
```

```
#2. 匹配单个字符(.)
patt314 = '3.14'
pi_patt = '3\.14'
m = re.match(pi_patt, '3.14')
if m is not None: print m.group()

m = re.match(patt314, '3014')
if m is not None: print m.group()
```

```
#3. 创建字符集合([])
m = re.match('[cr][23][dp][o2]', 'c3po')
if m is not None: print m.group()
```

```
#4. 重复、特殊字符和子组
patt = '\w+@(\w+\.)?\w+\.\com'
print re.match(patt, 'nobody@XXX.com').group()
print re.match(patt, 'nobody@www.xxx.com').group()
patt = '\w+@(\w+\.)*\w+\.\com'
print re.match(patt, 'nobody@www.xxx.yyy.com').group()
m = re.match('(\w{3})-(\d{3})', 'abc-123')
print m.group()      #所有匹配的部分,和 m.group(0)相同
print m.group(1)    #匹配的子组 1
print m.group(2)    #匹配的子组 2
print m.groups()    #所有匹配子组 ('abc', '123')
m = re.match('(a(b))', 'ab') #两个子组
print m.group(), m.group(1), m.group(2), m.groups()
```

```
#5. 从字符串的开头或结尾匹配及在单词边界上的匹配
re.search('^The', 'The end.')    #匹配
re.search('^The', 'end.The')     #不匹配
```

```
re.search(r'\bthe','bit thedog')    #匹配,单词左边界  
#r'\bthe'中的r表示原始字符串(raw strings),即\b会被理解成一个正则表达式的一个特殊符号,而不是退格键  
re.search(r'\bthe','bit the dog')   #匹配  
re.search(r'\bthe','bitthe dog')    #不匹配  
re.search(r'\Bthe','bitthe dog')   #匹配,单词右边界  
re.search(r'\Bthe','bit the dog')  #不匹配
```

#6. 用 `.findall()`找到每个出现的匹配部分

```
#结果: ['a','a','a']  
print re.findall('c(.)r', 'carry the barcardi to the car')  
#结果: ['car', 'car', 'car']  
print re.findall('c.r', 'carry the barcardi to the car')
```

#7. 用 `sub()`[和 `subn()`]进行搜索和替换

```
print re.sub('X','Mr. Smith', 'attn: X\n\nDear X,\n')  
#结果: ('XbcdXf', 2)  
print re.subn('[ae]', 'X', 'abcdef')
```

#8. 用 `split()`分割(分割模式)

```
# re模块和正则表达式的方法 split()与字符串的 split()方法相似,  
#前者根据正则表达式模式分割  
# 字符串,后者根据固定的字符串分割.  
#结果: ['123', '', '456', '789']  
print "123\n\n456\n789".split('\n')  
#结果: ['123', '456', '789']  
print re.split('\n+', "123\n\n456\n789")
```

#9. 贪心匹配

```
# 正则表达式本身默认是贪心匹配的。即,如果正则表达式模式中  
#使用到通配字,那它在按照从左到右  
# 的顺序求值时,会尽量抓取满足模式的最长字符串。  
testdata = 'gov::123143294538-6-8'  
patt = '.+(\d+-\d+-\d+)'  
print re.match(patt, testdata).group(1) #结果: 8-6-8  
#结果: 123143294538-6-8  
print re.search('\d+-\d+-\d+', testdata).group()  
# 解决贪心匹配的一个办法是用"非贪婪"操作符"?,"  
#这个操作符可以用在*,+,或?的后面.它的作用是要求  
# 正则表达式引擎匹配的字符越少越好。  
#结果: 123143294538-6-8  
print re.match('.+?(\d+-\d+-\d+)', testdata).group(1)
```

二十九、 Python 中的进程与线程

1) 进程和线程

进程和线程作为操作系统中的两个重要概念，即有相似之处，也有不同之处。本节首先介绍两者的概念，然后对 Python 的实现方式进行简要介绍。

与进程和线程相关的 Python 模块

模块名称	说明
os/sys	包含基本进程管理函数
subprocess	Python 基本库中多进程相关模块
signal	Python 基本库中信号相关模块
threading	Python 基本库中线程相关模块

2) 进程的运行环境

在 Python 中，os 模块提供了 `environ` 属性，用来记录当前进程的运行环境，这是一个字典型的数据结构，其中 `key` 为环境变量的变量名，`value` 为环境变量的值。

```
#-*-coding:utf-8-*-
import os
# 获取 PATH 的环境变量
path = os.environ.get('PATH')
print path
# 生成一个新的环境变量
os.environ['chuguangming'] = 'chuguangming'
# 获取所有的环境变量
for key in os.environ.keys():
    print key, '\t', os.environ[key]
```

3) 创建进程

创建进程是系统管理的重要组成部分，Python 语言提供了多种方式来创建进程，除了在 os 模块中有丰富的创建进程函数外，还有其他专门的模块如 subprocess 等来管理进程。在本节中主要对 os 模块中的进程创建函数进行说明。

在 os 模块中主要包含的 `system` 和 `exec` 家族函数，能够适应不同的创建进程需求，具体的将在下面进行介绍。同时，还有一些其他的创建进程方法，如 `os.popen` 等函数。

1. system 函数

`system` 函数是用来创建进程的最快捷的方式，此函数在新进程中执行 `command` 字符串命令。如果返回值为 0，表示命令执行成功，否则表示失败。现在，可以通过 `python` 来执行系统命令，如下代码：

```
#-*-coding:utf-8-*-
import os
print os.system('dir')
```

结果将输出这段代码所在目录的文件和文件夹。注意到程序最后还打钱出 0，表示这次命令执行成功。

三十、 Pyexe

有的时候大部分人对于 python 这种脚本化的运行方式非常的不爽，其实解决方式很多，今天给出一种方法。
PyInstaller. PyInstaller 可以将 Python 程序打包成 Windows (当然也包括 Linux, Mac OS X, Solaris and AIX) 下可执行的 EXE 文件，目前支持 python2.2-2.7 版本，点击这里下载。

使用 PyInstaller 需要安装 PyWin32，可到这里下载相应的版本。(从 pywin32 的下载量看，还是 Python2.7 使用更广泛)

PyInstaller 下载地址: <http://www.pyinstaller.org/>

PyWin32 下载地址: <http://sourceforge.net/projects/pywin32/files/pywin32/Build%20217/>

pywin32-217.win32-py3.3.exe	2012-02-12	6.6 MB		
pywin32-217.win32-py3.2.exe	2012-02-12	6.6 MB		
pywin32-217.win32-py3.1.exe	2012-02-12	6.6 MB		
pywin32-217.win32-py2.7.exe	2012-02-12	6.6 MB		
pywin32-217.win32-py2.6.exe	2012-02-12	6.6 MB		
pywin32-217.win32-py2.5.exe	2012-02-12	5.6 MB		
pywin32-217.win32-py2.4.exe	2012-02-12	5.6 MB		
pywin32-217.win32-py2.3.exe	2012-02-12	5.0 MB		

- 1、下载对应已安装的 Python 版本的 PyInstaller 版本，解压到任意目录，按照提供的 manual 文档进行即可
- 2、假设要转换的 Python 代码为

C:\PythonProject\testurllib\urllib2test1.py

- 3、进入 PyInstaller 安装根目录，执行以下命令。

python PyInstaller.py C:\PythonProject\testurllib\urllib2test1.py

- 4、结果

名称	修改日期	类型	大小
hashlib.pyd	2012/4/10 23:30	PYD 文件	461 KB
bz2.pyd	2012/4/10 23:24	PYD 文件	79 KB
Microsoft.VC90.CRT.manifest	2013/3/5 13:10	MANIFEST 文件	2 KB
msxml3.dll	2013/3/5 13:10	应用程序扩展	240 KB
msxml6.dll	2013/3/5 13:10	应用程序扩展	832 KB
msasn1.dll	2013/3/5 13:10	应用程序扩展	613 KB
python27.dll	2012/4/10 23:24	应用程序扩展	2,918 KB
select.pyd	2012/4/10 23:24	PYD 文件	11 KB
unicodedata.pyd	2012/4/10 23:24	PYD 文件	674 KB
urllib2test1.exe	2013/4/25 14:03	应用程序	830 KB
urllib2test1.exe.manifest	2013/4/25 14:03	MANIFEST 文件	1 KB

常用参数：

-F 制作独立的可执行程序
-D 制作出的档案存放在同一个文件夹下（默认值）
-K 包含 TCL/TK（对于使用了 TK 的，最好加上这个选项，否则在未安装 TK 的电脑上无法运行）
-W 制作窗口程序
-C 制作命令行程序（默认）
-X 制作使用 UPX 压缩过的可执行程序

UPX 下载地址：<http://upx.sourceforge.net/download/upx308w.zip>

-o DIR 指定输出 SPEC 文件路径（这也决定了最后输出的 exe 文件路径）
--icon=[ICO 文件路径] 指定程序图标
-v [指定文件] 指定程序版本信息
-n [指定程序名] 指定程序名称

具体参数见：安装文档，非常的丰富

三十一、 Sina SDK 的使用

1) 首先申请 SDK 开发者

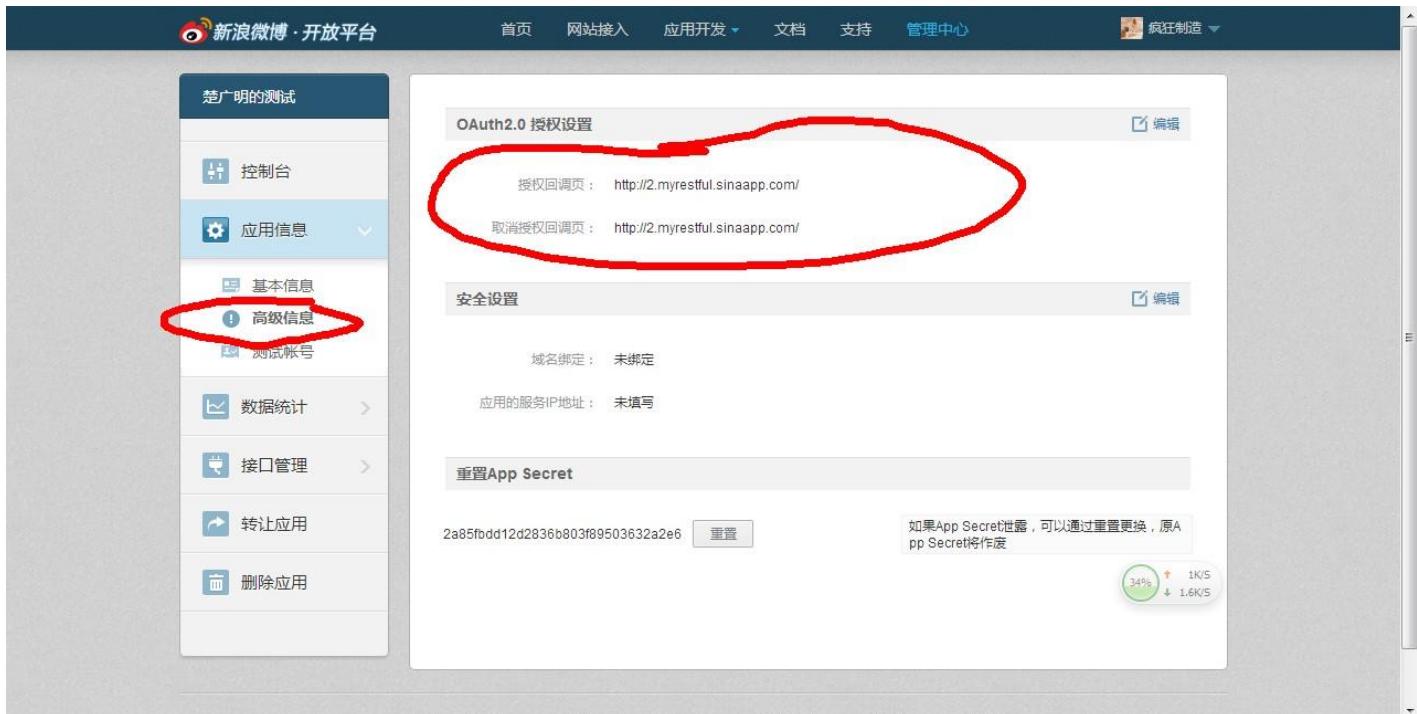
1、首先访问 <http://open.weibo.com/> 申请开发者

楚广明新员工培训专用（借用了无数开源先驱的文档与资源，有些出处以无法考证。向开源文档的先驱致敬！）

The screenshot shows the Weibo Open Platform homepage. At the top, there's a banner about migration to V1 interface. Below it, a large graphic of a tree growing from a city skyline on a hill. A button labeled '微博认证团队' (Weibo Certified Team) is visible. Three main service categories are highlighted: '网站接入' (Website Access), '移动应用' (Mobile Application), and '站内应用' (Internal Application). Below these are sections for '成功案例' (Success Stories) and '平台新动态' (Platform News). Success stories include hoopCHINA, 易车网, M18.com, and others. Platform news items mention the opening of disaster area data and a timestamp of February 26, 2013.

The screenshot shows the Weibo Management Center. It features a sidebar with user information and payment account details. The main area is titled '管理中心' (Management Center) and includes tabs for '我的应用' (My Applications) and '我的网站' (My Websites). A red arrow points to a button labeled '创建新的应用' (Create a new application). Another red arrow points to a message indicating '未提交审核' (Not submitted for review) and '需完成：完善应用信息' (Requires completion: Improve application information). A progress bar at the bottom right shows 34% completion.

在申请的时候其中很简单，大家可以参考 SDK 的帮助，唯一要注意的是，回调地址一定要真实，要不然以后会有麻烦的。



2) 下载 Python SDK

去下载 SDK，地址：<http://michaelliao.github.io/sinaweibopy/>

3) 一个简单的演示程序

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-
from weibo import APIClient
import sys
import webbrowser

APP_KEY = '1705534647'      # 填写申请的 APP_KEY
APP_SECRET = '2a85fbdd12d2836b803f89503632a2e6'      # 填写申请的 APP_SECRET
CALLBACK_URL = 'http://2.myrestful.sinaapp.com/'

''' posttext.py 调用 weibo SDK 命令发微博 '''

# 获取微博授权，手动操作

def getclient():
    client = APIClient(app_key=APP_KEY, app_secret=APP_SECRET,
    redirect_uri=CALLBACK_URL)
```

```
url = client.get_authorize_url()

print url    # 浏览器打开该 url, 取得 code='xxx' 类似的 code 输入
# 将地址保存在文件中, 可以手工的打开这个文件, 在地址栏中显示
# 之后转向到一个新的 url
#
#https://api.weibo.com/oauth2/default.html?code=96296f963705cbca77b673ac7fbdf9d0
# 其中的 code=后面就是 code
f = file("url.txt", "w+")
f.writelines(url)
f.close()

sinaauthurl = urllib.urlopen(url)
#print sinaauthurl.read().decode('utf-8')
webbrowser.open_new_tab(url)

code = raw_input('Enter code >')

r = client.request_access_token(code)
access_token = r.access_token
expires_in = r.expires_in

client.set_access_token(access_token, expires_in)

return client

# 发微博

def posttext(client):
    text = raw_input('Enter text to post >')
    utext = unicode(text, "UTF-8")
    client.statuses.update.post(status=utext)

if __name__ == '__main__':
    client = getclient()
    posttext(client)
```

4) 一个进阶的演示

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-
from weibo import APIClient
import sys
import webbrowser
```

```
APP_KEY = '1705534647'      # 填写申请的 APP_KEY
APP_SECRET = '2a85fbdd12d2836b803f89503632a2e6'      # 填写申请的 APP_SECRET
CALLBACK_URL = 'http://2.myrestful.sinaapp.com/'

''' posttext.py 调用 weibo SDK 命令发微博 '''

# 获取微博授权，手动操作

def getclient():
    client = APIClient(app_key=APP_KEY, app_secret=APP_SECRET, redirect_uri=CALLBACK_URL)
    url = client.get_authorize_url()

    print url      # 浏览器打开该 url，取得 code='xxx' 类似的 code 输入
    # 将地址保存在文件中，可以手工的打开这个文件，在地址栏中显示
    # 之后转向到一个新的 url
    # https://api.weibo.com/oauth2/default.html?code=96296f963705cbca77b673ac7fbdf9d0
    # 其中的 code=后面就是 code
    f = file("url.txt","w+")
    f.writelines(url)
    f.close()
    # sinaauthurl = urllib.urlopen(url)
    # print sinaauthurl.read().decode('utf-8')
    webbrowser.open_new_tab(url)

    code = raw_input('Enter code >')

    r = client.request_access_token(code)
    access_token = r.access_token
    expires_in = r.expires_in

    client.set_access_token(access_token, expires_in)

    return client

# 发微博

def posttext(client):
    text = raw_input('Enter text to post >')
    utext = unicode(text, "UTF-8")
    client.statuses.update.post(status=utext)

#绝大部分都需要获取用户信息
def getmyinfo(client):
    r = client.users.show.get(screen_name='疯狂制造')
    return r

def getmyfanscount(client,id):
```

```
r=client.users.counts.get(uids=id)
return r
if __name__ == '__main__':
    client = getclient()

    allfans= getmyinfo(client)
    #得到用户 ID
    print allfans.id
    print allfans.description

    #得到用户有多少 fans
    print getmyfanscount(client,allfans.id)
    # posttext(client)
```

三十二、 附录-Ubuntu 开发环境之安装

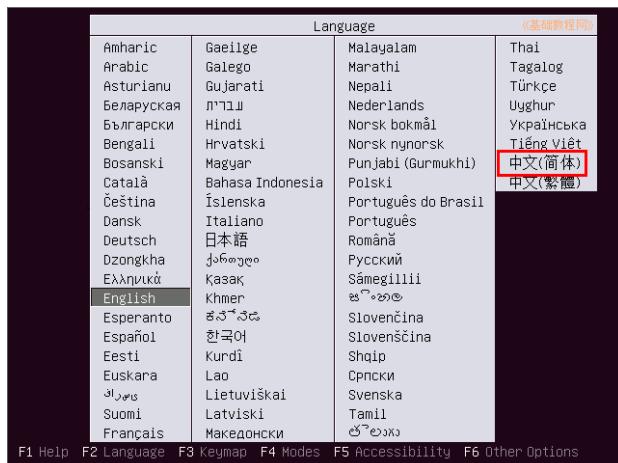
内容来之于 <http://teliute.org>, 非常好的新手教程网站

1) 进入 live cd 桌面

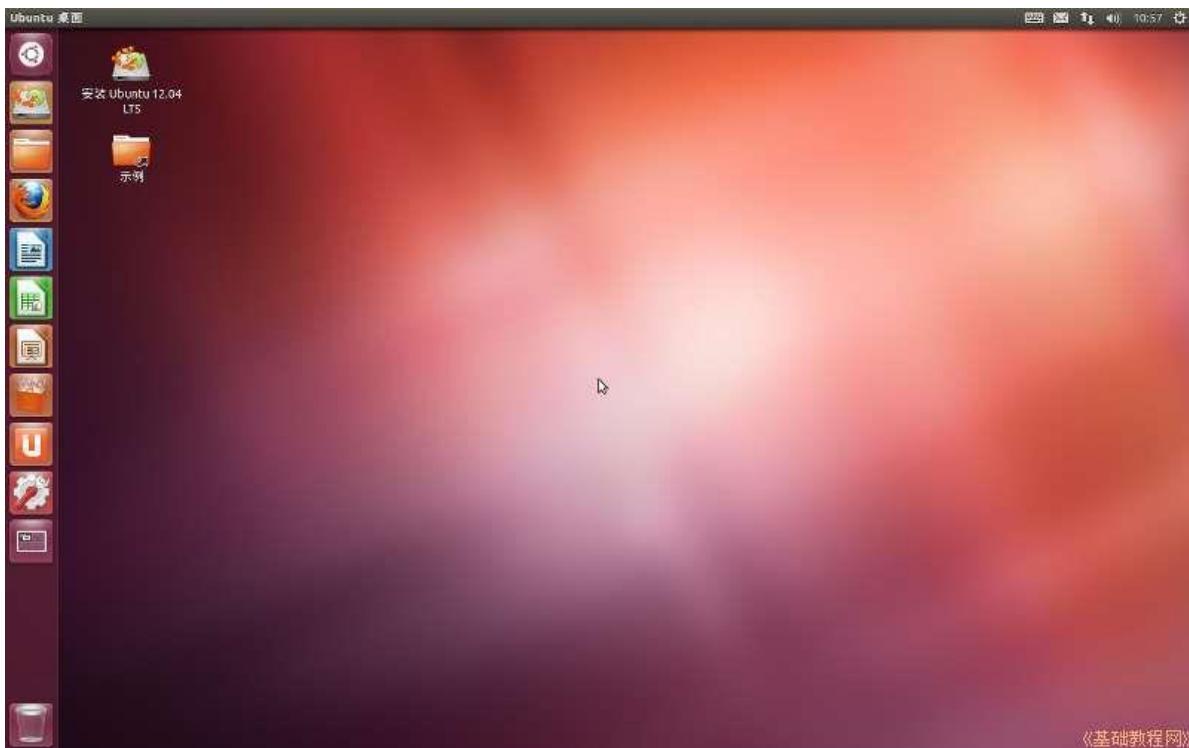
1) 设置好[启动](#)后, 断开网络, 然后重启动计算机, 可以用硬盘启动, 也可以刻成光盘启动, 镜像的下载地址:
进入后找蓝色链接点击下载, 如 `ubuntu-12.04-desktop-i386.iso`, 64 位 CPU 可以下载 `amd64` 的版本:

<http://mirrors.sohu.com/ubuntu-releases/12.04/>

2) 启动后稍等, 系统自动运行, 在下边出来两个图标时  , 可以按一下 `ESC` 键呼出菜单项, 再按右方向键选择“中文(简体)”, 按回车键;



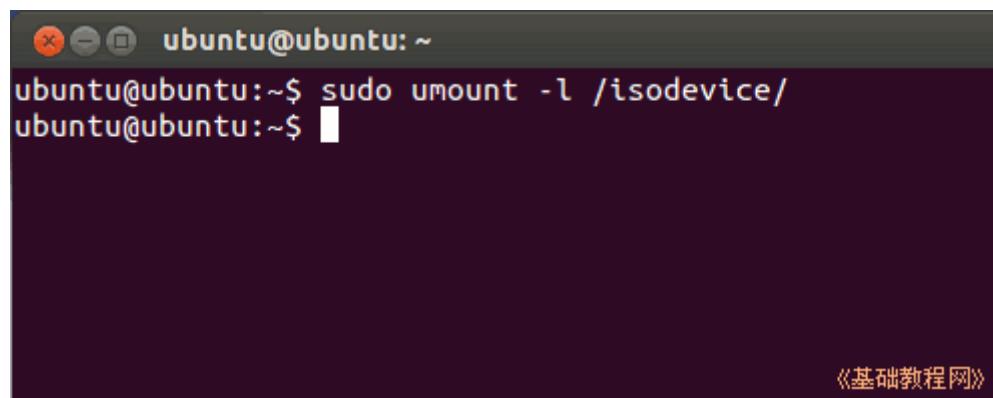
3) 等一会就进入一个桌面，这就是试用的 `live cd` 桌面，**桌面**左上边有有两个图标，右上角是“**关机**”按钮；



4) 对于**硬盘**安装，点一下左上角的圆圈按钮，稍等在旁边出来的文本框中输入字母 `ter` 然后点击下边出来的终端图标；



5) 输入命令 `sudo umount -l /isodevice` 然后按一下**回车键**，没什么提示就是成功了，关闭终端；



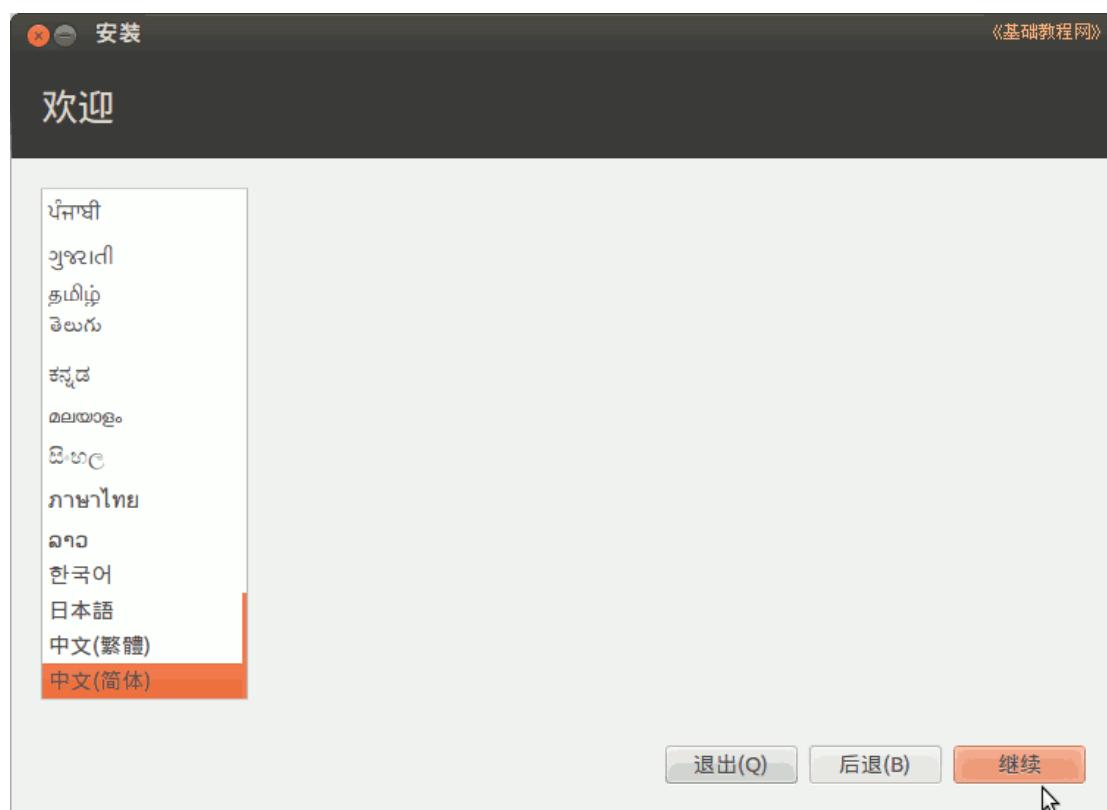
```
ubuntu@ubuntu:~$ sudo umount -l /isodevice/
ubuntu@ubuntu:~$ █
```

《基础教程网》

2) 安装系统

1) 双击桌面“**安装 Ubuntu12.04 LTS**”图标，稍等出来一个“欢迎”面板，左侧应该选中了“中文(简体)”，

如果不是就在左边选中它，然后点右下角“**继续**”按钮；



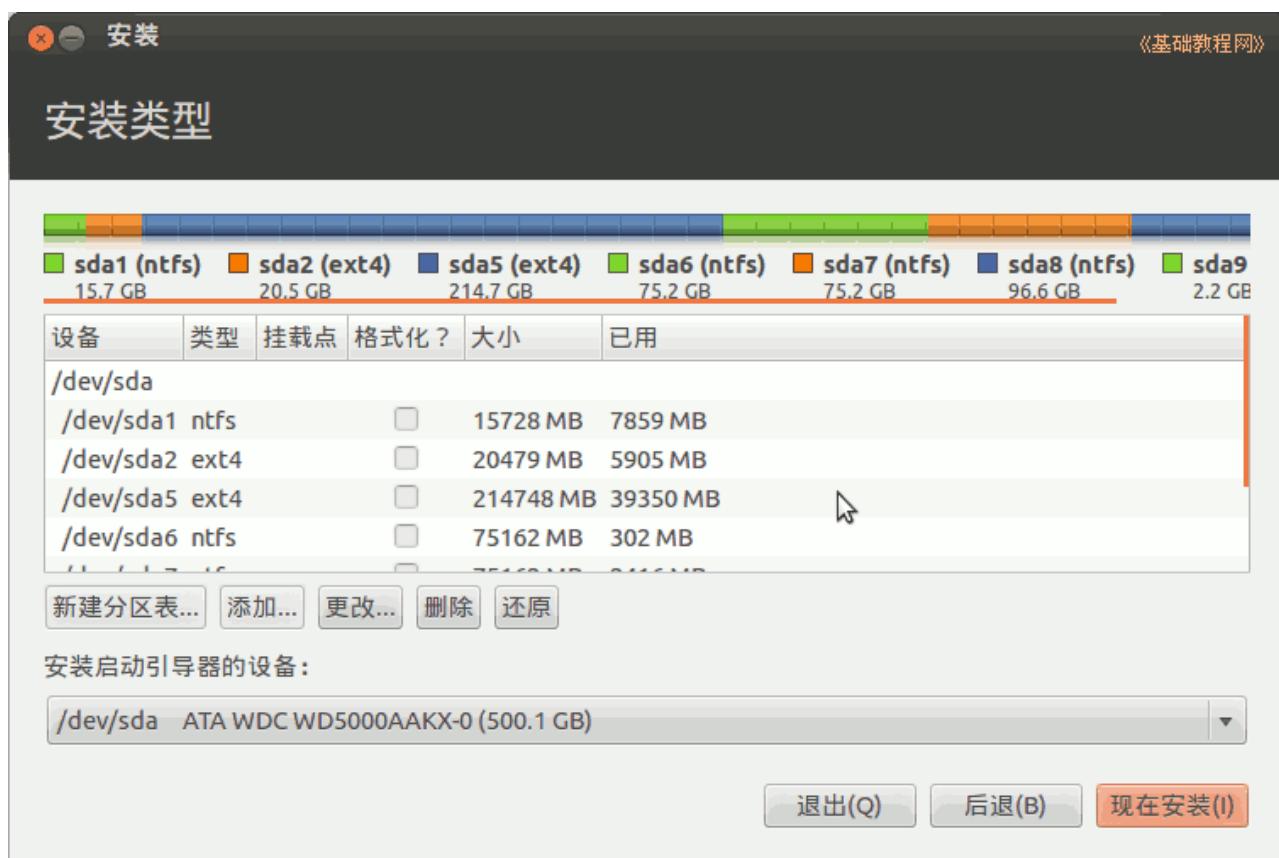
2) 第 2 步是检查准备情况，要求磁盘空间足够，不要连接网络，一般不勾选更新和第三方软件，**直接**点“继续”按钮；



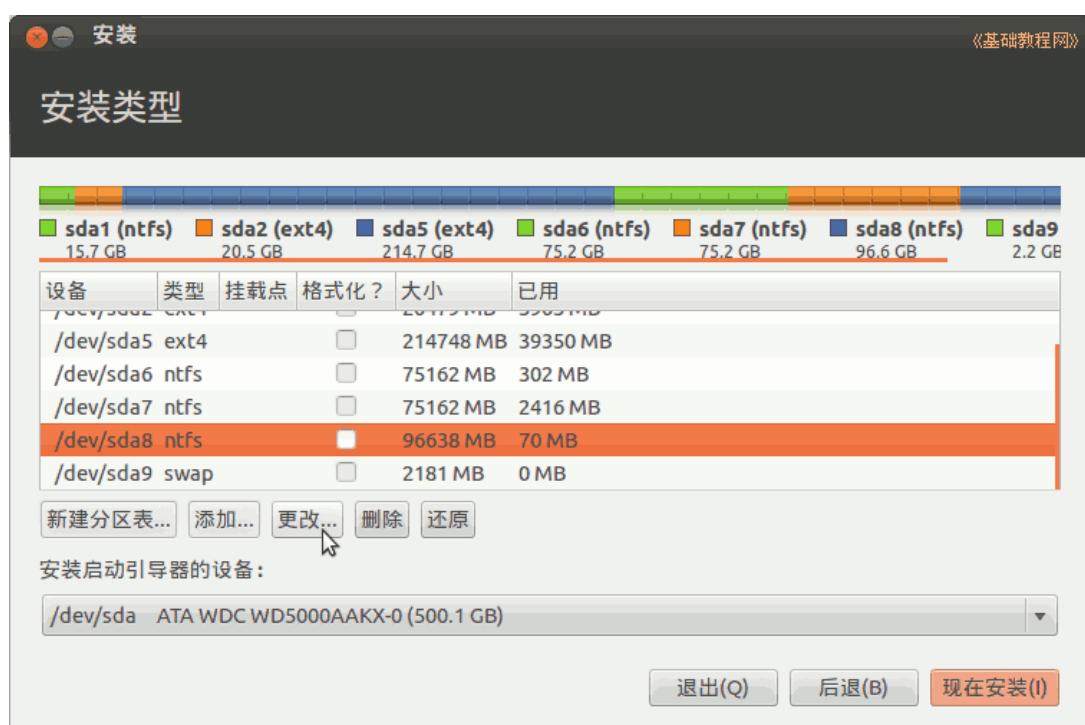
3) 接下来**第3步**是询问安装到哪个分区，选择最下边的“其他选项”，点“**继续**”按钮；



- 4) 接下来出来磁盘分区情况，如果要新建分区和转换分区可以参考前面的第 7、13、15、16、19 课，这儿是**安装**到一个 NTFS 分区上；



- 5) 点击选中计划要安装的分区，可以根据分区类型和大小来确定，然后点下边的“**更改**”按钮；

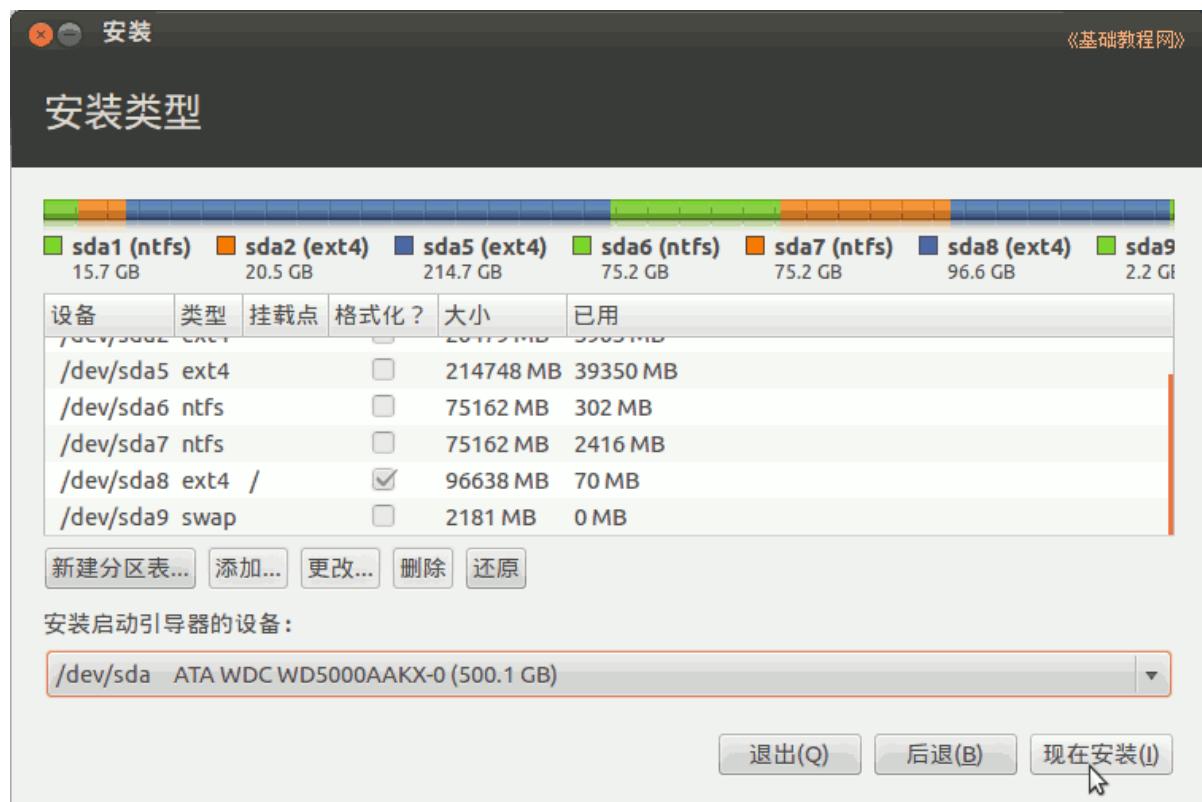


6) 在出来的对话框中，设定**用于**分区的格式 Ext4，打勾“格式化”，在“挂载点”右边点一下，选“/”，点“确定”

注意，格式化会删除这个分区上的所有文件，请提前备份重要数据；



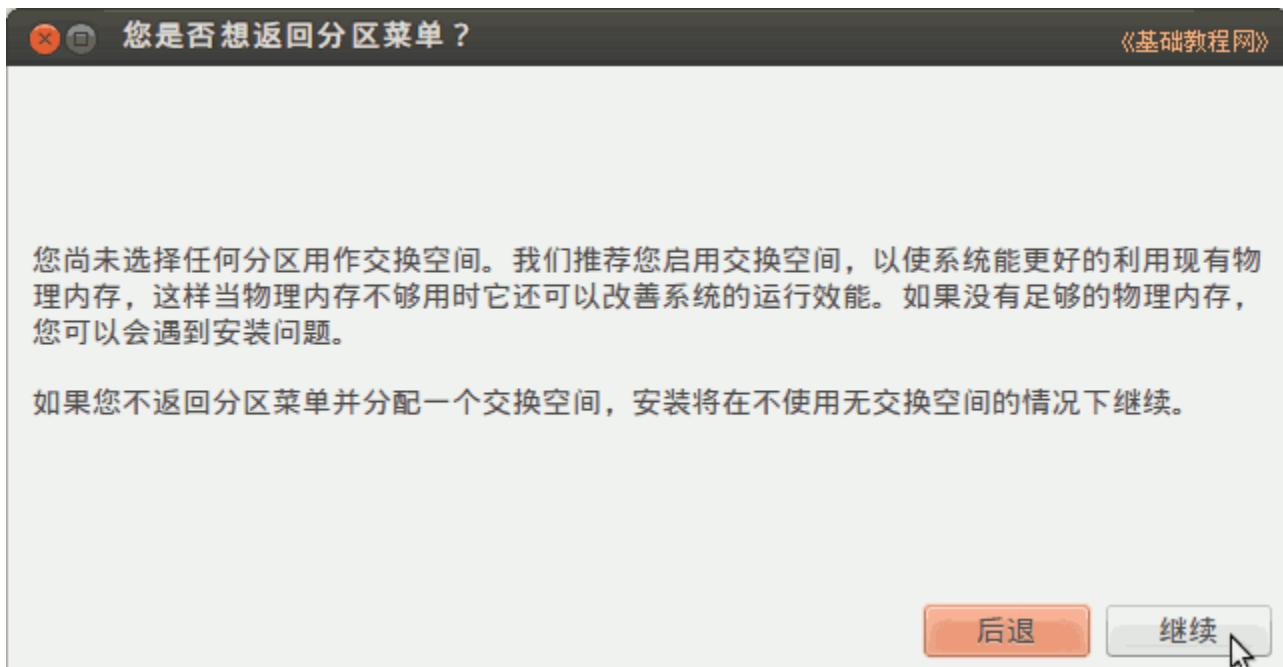
7) 回到分区面板，检查一下**分区**编辑好了，点“现在安装”，如果还有 /home 分区，按原来的设，一般不打勾“**格式化**”，只需提前清理里面的配置文件；



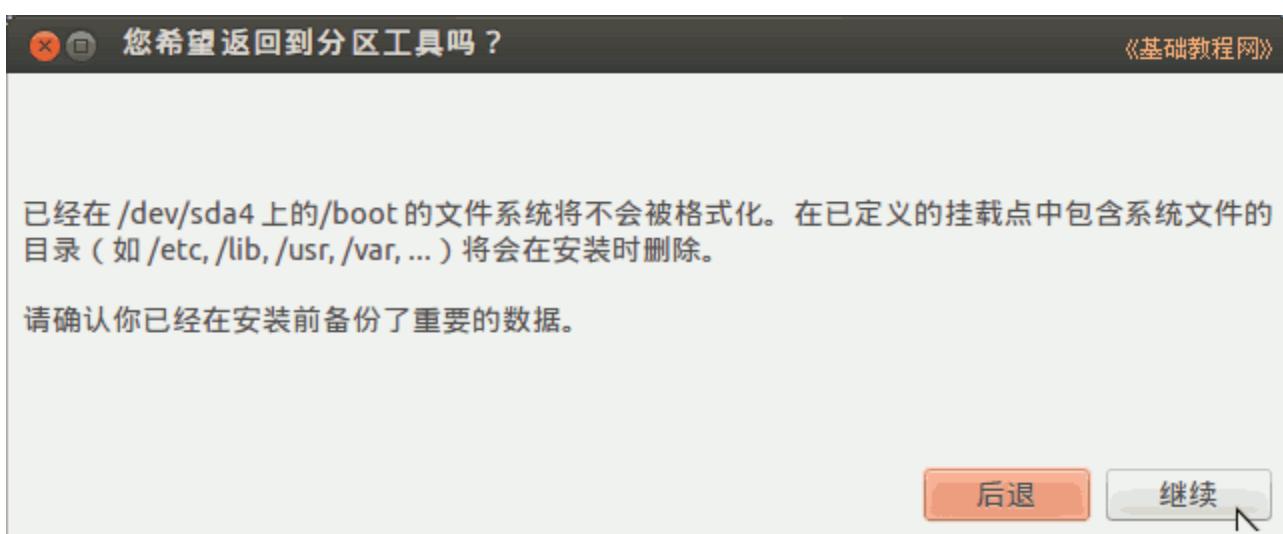
8) 在下边是安装 grub 引导器的选项，可以选择安装到/分区，然后用 Windows 来引导 Ubuntu，参阅：[用](#)

EasyBCD 在 Win 7/8 硬盘安装 Ubuntu

如果提示没有交换空间，点“继续”，这个是用于休眠的，大小跟内存相同的分区，一般不用它；



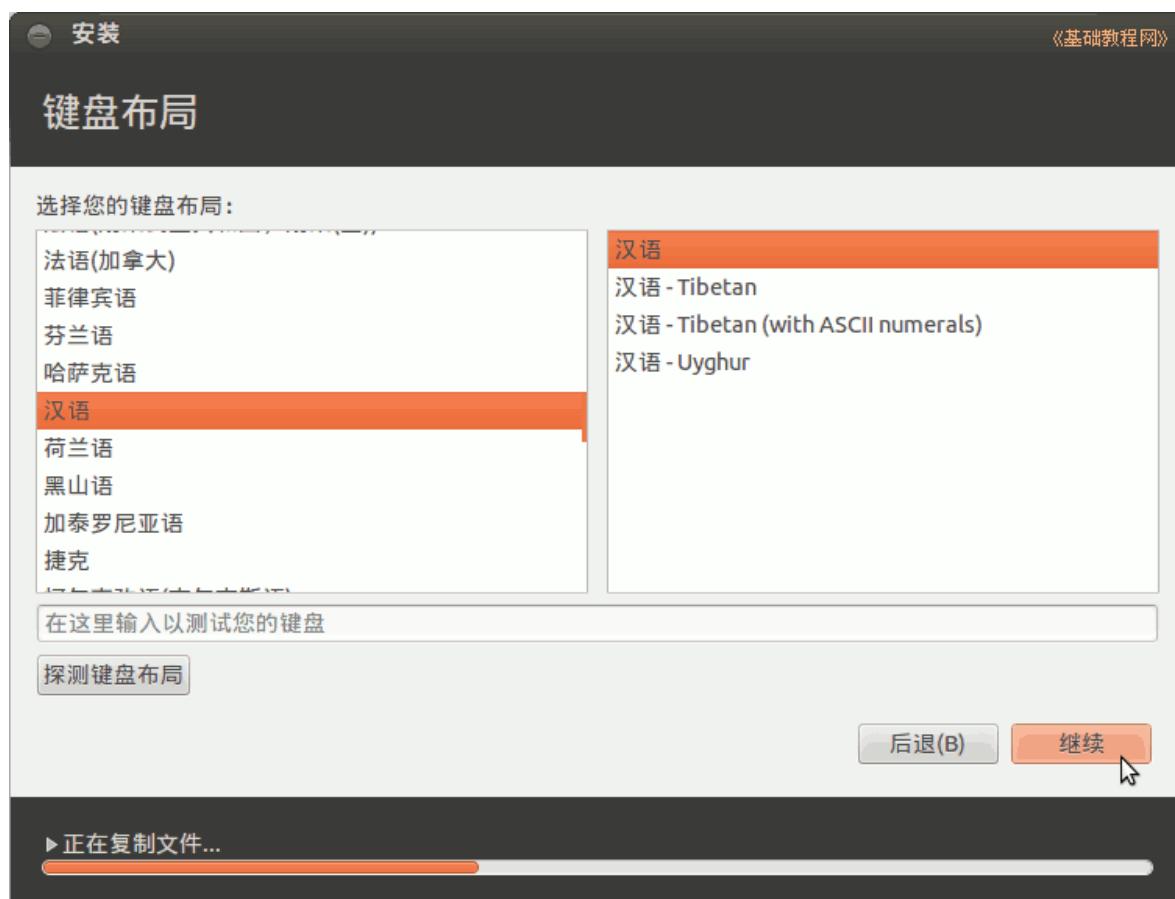
提示即将删除和格式化时，确保重要数据已经备份，点“继续”；



9) 然后出来询问地区，点“继续”即可，或者在地图上找找自己所在的地区，这时下边已经开始安装了；



10) 接下来是键盘布局，一般是“USA”，直接点“前进”继续；



11) 接下来是设定自己的用户名(小写字母)和密码等，从上到下依次输入即可，然后点“继续”；



12) 然后接着继续安装过程，可以看一下系统的介绍；



13) 耐心等待完成，然后出来一个对话框，点“**现在重启**”完成安装，按电源键也可以呼出关机对话框；



14) 如果是光盘安装，稍等提示取出光盘，然后按**回车键**，重新启动计算机，安装完成；

3) 连网换源

1) 重新启动后，停在一个**登录**界面，点击自己的用户名，输入密码后按回车，进入系统；



2) 进入桌面后会出来一个“**不完整**语言支持”的提示对话框，先不关闭拖到一边后面要用；



3) 先连网 , 找到屏幕**右上角**键盘旁边的一个网络图标 , 瞄准点**左键** , 选 “编辑连接” ;



4) 对于 ADSL 拨号宽带连接上网 , 在出来的对话框右边 , 点 **DSL** 标签 , 然后点添加 ;



5) 在出来的对话框里 , 先**打勾**上边的 “自动连接” , 然后在用户名里输入 **ADSL 宽带用户名** ,

下面的密码里 , 输入宽带**密码** , 然后点右下角的 “应用” 按钮 , 在出来的密码框中输入密码 ;



关闭对话框以后，稍等一下会提示连网成功，



也可以**重新**启动计算机，命令行使用 `sudo pppoeconf` 命令也可以；

(下面 6-7 是局域网的，宽带用户不要设置)

6) 如果是**局域网**上网，在出来的对话框里，选中里面的“`Wired connection1`”，点右边的“编辑”按钮；



7) 在出来的对话框里，选择 **IPV4 设置** 标签，在下面的方法里选“手动”，再点下边的“添加”按钮，在出来的文本框里，**依次**输入 IP 地址、子网掩码、网关，然后按**回车键**确定，在下面的 DNS 里输入 DNS 服务器地址，检查一下，点右下角的“**保存**”按钮，在出来的认证中，输入自己的密码；



回到原来对话框点“关闭”，稍等一会就可以上网了，也可以重新启动计算机；

8) 连接好网络后先**换源**，在左边的快速启动面板中点击“软件中心”按钮，再把鼠标移到最上边出现菜单栏，点“编辑—软件源”；



9) 在出来的软件源面板下边，点击“中国的服务器”下拉列表，选择“其他站点”；



10) 在出来的服务器列表中，选择搜狐、163 或 cn99 的站点都可以，然后点右下角的“选择服务器”按钮返回；



11) 关闭窗口后，出来身份验证，输入自己的登录密码后点“授权”，如果出来更新提示，点“重新载入”，等待完成就好了；



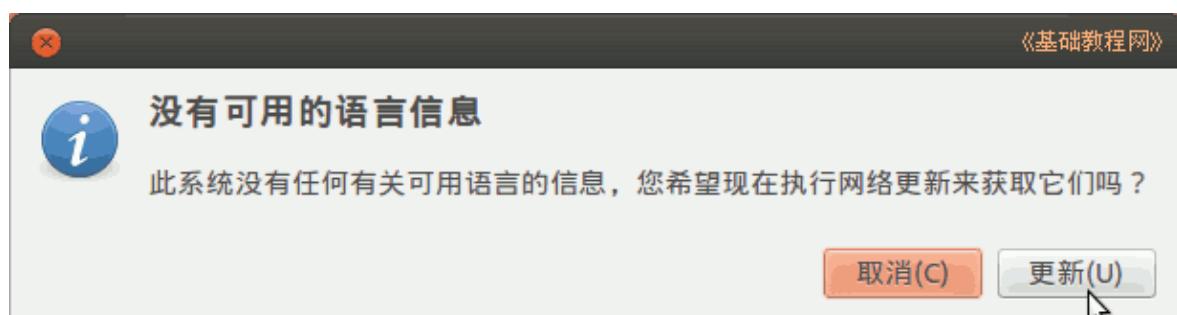
更多设置源的相关介绍，请参阅：<http://teliute.org/linux/Ubsetup/lesson9/lesson9.html>

4) 更新系统

1) 回到最开头的“不完整语言支持”对话框，点“现在执行此动作”，如果找不着了，就点右上角齿轮图标—系统设置—语言支持；



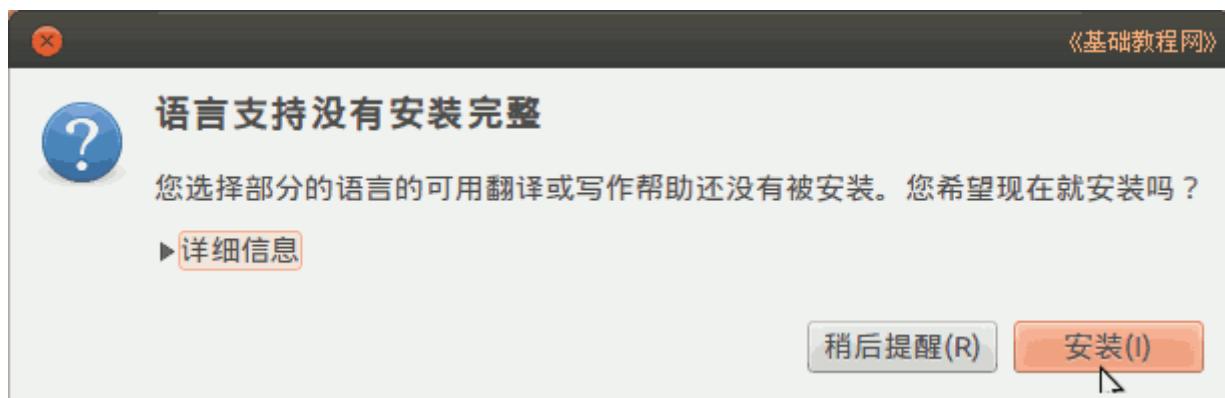
2) 如果出来提示“没有可用的语言信息”面板，点击“更新”按钮；



3) 然后出来一个缓存更新的对话框，耐心等待完成，点“详情”可以查看进度；



4) 然后在出来的“语言支持没有完全安装”对话框中，点“安装”，安装语言包；



5) 待语言包安装完成后，点**关闭**按钮关闭对话框；



6) 在左边的快速启动面板中，还会有一个更新按钮(上边标有数字)，点击出来更新管理器窗口，或者点右键选“更新管理器”；



7) 在出来的窗口右下角, 点击“安装更新”按钮, 左边的“设置”里也可以设置软件源;



8) 接下来就是下载和安装，这儿**耐心**等待即可；



9) 安装完成后，点右上角的“现在重启”按钮重启电脑，然后就可以开始使用了；



其他细节：如果发现 Windows 的启动菜单项没了，先不着急，更新完就有了，如果还没有则在终端执行 `sudo update-grub`，或者把原来备份的 `grub.cfg` 找出来，照着填加进去；

三十三、附录 vmware 中的 bridge、nat 区别

概述：

VMWare 提供了三种工作模式，它们是 **bridged**(桥接模式)、**NAT**(网络地址转换模式)和 **host-only**(主机模式)。要想在网络管理和维护中合理应用它们，你就应该先了解一下这三种工作模式。

1.bridged(桥接模式)

在这种模式下，VMWare 虚拟出来的操作系统就像是局域网中的一台独立的主机，它可以访问网内任何一台机器。

在桥接模式下，你需要手工为虚拟系统配置 IP 地址、子网掩码，而且还要和宿主机器处于同一网段，这样虚拟系统才能和宿主机器进行通信。同时，由于这个虚拟系统是局域网中的一个独立的主机系统，那么就可以手工配置它的 TCP/IP 配置信息，以实现通过局域网的网关或路由器访问互联网。

使用桥接模式的虚拟系统和宿主机器的关系，就像连接在同一个 Hub 上的两台电脑。想让它们相互通信，你就需要为虚拟系统配置 IP 地址和子网掩码，否则就无法通信。

如果你想利用 VMWare 在局域网内新建一个虚拟服务器，为局域网用户提供网络服务，就应该选择桥接模式。

2.host-only(主机模式)

在某些特殊的网络调试环境中，要求将真实环境和虚拟环境隔离开，这时你就可采用 host-only 模式。在 host-only 模式中，所有的虚拟系统是可以相互通信的，但虚拟系统和真实的网络是被隔离开的。

提示：在 host-only 模式下，虚拟系统和宿主机器系统是可以相互通信的，相当于这两台机器通过双绞线互连。

在 host-only 模式下，虚拟系统的 TCP/IP 配置信息(如 IP 地址、网关地址、DNS 服务器等)，都是由 VMnet1(host-only)虚拟网络的 DHCP 服务器来动态分配的。

如果你想利用 VMWare 创建一个与网内其他机器相隔离的虚拟系统，进行某些特殊的网络调试工作，可以选择 host-only 模式。

3.NAT(网络地址转换模式)

使用 NAT 模式，就是让虚拟系统借助 NAT(网络地址转换)功能，通过宿主机器所在的网络来访问公网。也就是说，使用 NAT 模式可以实现在虚拟系统里访问互联网。NAT 模式下的虚拟系统的 TCP/IP 配置信息是由 VMnet8(NAT)虚拟网络的 DHCP 服务器提供的，无法进行手工修改，因此虚拟系统也就无法和本局域网中的其他真实主机进行通讯。采用 NAT 模式最大的优势是虚拟系统接入互联网非常简单，你不需要进行任何其他的配置，只需要宿主机器能访问互联网即可。

如果你想利用 VMWare 安装一个新的虚拟系统，在虚拟系统中不用进行任何手工配置就能直接访问互联网，建议你采用 NAT 模式。

提示：以上所提到的 NAT 模式下的 VMnet8 虚拟网络，host-only 模式下的 VMnet1 虚拟网络，以及 bridged 模式下的 VMnet0 虚拟网络，都是由 VMWare 虚拟机自动配置而生成的，不需要用户自行设置。VMnet8 和 VMnet1 提供 DHCP 服务，VMnet0 虚拟网络则不提供。

这个问题一直都没有彻底的搞清楚，心里不是很很有数，Vmware 中什么网桥，NAT 啊，Hostonly 什么的，另外还有 Colinux 里面的网络问题。总之很繁杂，搞清楚很有必要。

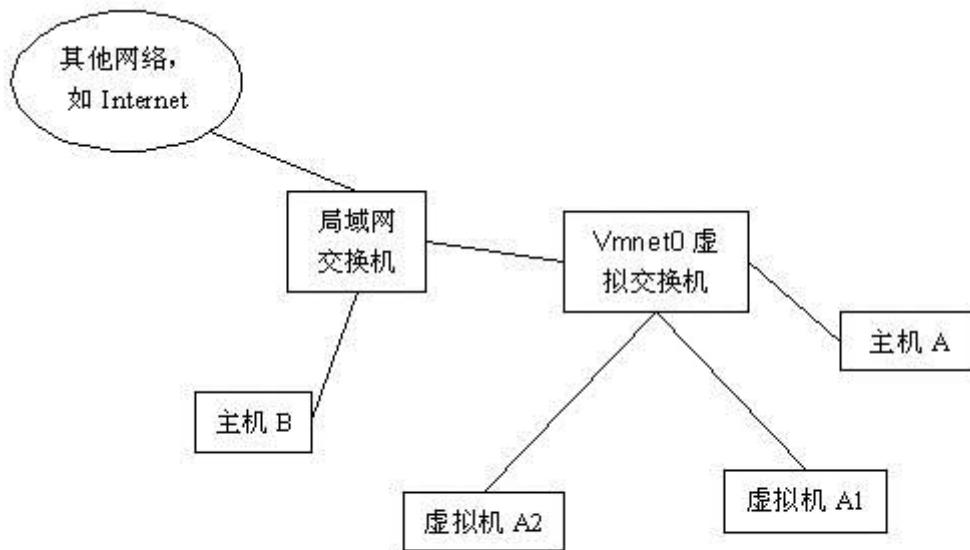
1) VMware 网络连接的几种方式

大家在安装完虚拟机后，默认安装了两个虚拟网卡，VMnet1 和 VMnet8，其他的未安装（当然也可以手动安装其他的）。其中 VMnet1 是 host 网卡，用于 host 方式连接网络的。VMnet8 是 NAT 网卡，用于 NAT 方式连接网络的。它们的 IP 地址是随机生成的，如果要用虚拟机做实验的话，最好将 VMnet1 到 VMnet8 的 IP 地址改掉。习惯上把 VMware 虚拟网卡使用的网段“固定”，使用如下原则：VMnet1 对应的网段是 192.168.10.0，VMnet2 对应的网段是 192.168.20.0，其他的类似。当然平常只是随便用用的就不用改了，能上网就行了。

网络连接的方式主要有：Bridged（桥接）、NAT、host（主机网络）。

Use bridged networking（使用桥接网络）

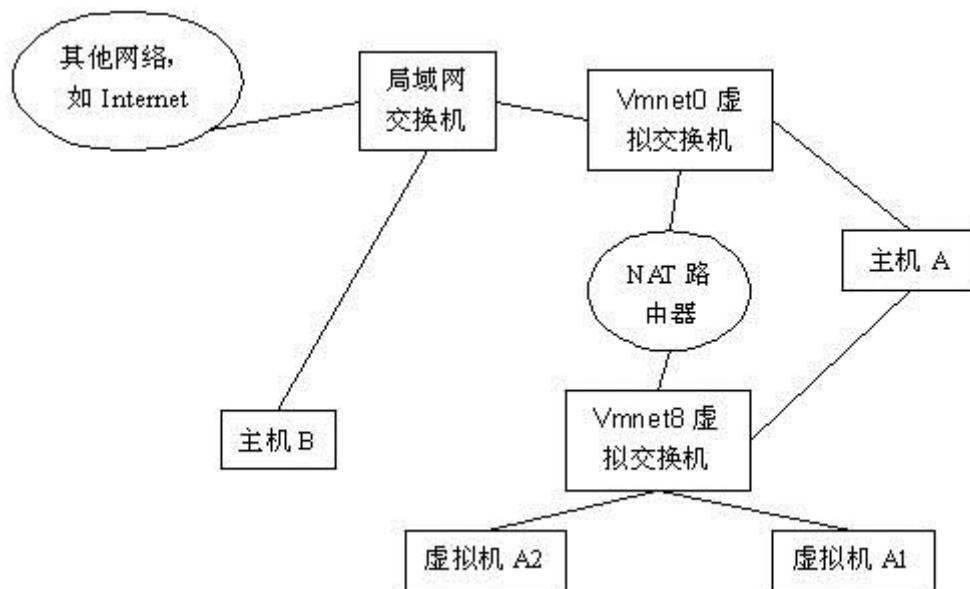
说明：使用 `Vmnet0` 虚拟交换机，此时虚拟机相当与网络上的一台独立计算机与主机一样，拥有一个独立的 IP 地址。



使用桥接方式，A, A1, A2, B 可互访。

Use network address translation (NAT)

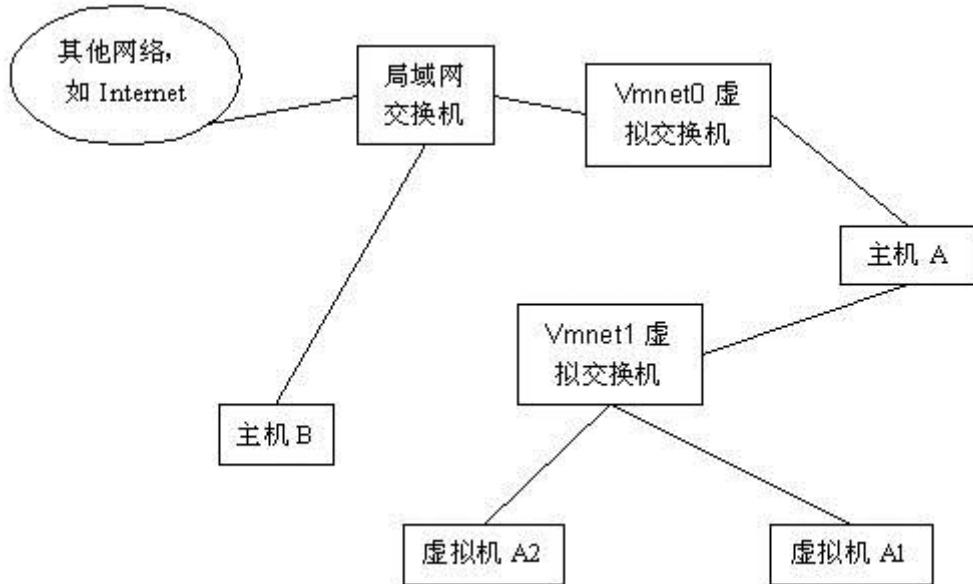
说明：使用 `Vmnet8` 虚拟交换机，此时虚拟机可以通过主机单向网络上的其他工作站，其他工作站不能访问虚拟机。



使用 NAT 方式，A1, A2 可以访问 B，但 B 不可以访问 A1, A2。但 A, A1, A2 可以互访。

Use Host-Only networking（使用主机网络）

说明：使用 `Vmnet1` 虚拟交换机，此时虚拟机只能与虚拟机、主机互访。也就是不能上 `Internet`



使用 Host 方式，A, A1, A2 可以互访，但 A1, A2 不能访问 B，也不能被 B 访问。

实际应用：现在最普遍使用的是宽带，使用桥接方式，主机必然可以上网，但是虚拟机上的操作系统是不可以上网的。此时你的主机就相当于 A，而虚拟机就相当于 A1，除非你再向运营商（如电信，网通）申请个 IP 地址，显然这是不可取的。如果你有路由器的话，那么就可以使用桥接方式上网，因为此时你申请的 IP 地址是写在了路由器上，而不是你的机器上，这样包括你的主机，虚拟机，也包括连在路由器上的其他人的机器，都将可以上网，使用着由路由器自动分配的 IP 地址，IP 地址一般是 192.168.0.XXX。貌似运营商不太乐意你使用路由器，说什么对他们交换机的接口有损害。当然，只要你的主机可以上网，使用 NAT 方式，虚拟机总是可以上网（当然不包括在你胡乱改写 IP，网关，DNS 等信息的情况下）。

实例手动设置：假设主机的 VMnet1 使用网段地址为 192.168.10.0, VMnet8 使用网段地址为 192.168.80.0, 网关地址为 192.168.80.254（也就是 NAT 路由器地址，可自己设置，在 VMnet8 上），主机网卡使用地址为 192.168.1.1。

一、使用桥接方式，虚拟机 A1 的 IP 地址可以设置成 192.168.1.5（与主机网卡地址同网段的即可），其他的诸如网关地址，DNS，子网掩码均与主机的相同。

二、使用 NAT 方式，虚拟机 A1 的 IP 地址可以设置成 192.168.80.5（与 VMnet8 使用相同的网段），网关是 NAT 路由器地址，即 192.168.80.524

三、使用 host-only 方式，虚拟机 A1 的 IP 地址可以设置成 192.168.10.5（与 VMnet1 使用相同的网段）

说明：一般用户不需要自己手动设置，选择自动获取，或开启 DHCP 服务即可。

2. 第二套说法

很多朋友都用 vmware 来测试不同的系统，我结合自己的经验谈一下对网络设置的理解，不对的地方请指正。

bridge:

这种方式最简单，直接将虚拟网卡桥接到一个物理网卡上面，和 linux 下一个网卡 绑定两个不同地址类似，实际上是将网卡设置为混杂模式，从而达到侦听多个 IP 的能力。

在此种模式下，虚拟机内部的网卡（例如 linux 下的 eth0）直接连到了物理网卡所在的网络上，可以想象为虚拟机和 host 机处于对等的地位，在网络关系上是平等的，没有谁在谁后面的问题。

使用这种方式很简单，前提是你可以得到 1 个以上的地址。对于想进行种种网络实验的朋友 不太适合，因为你无法对虚拟机的网络进行控制，它直接出去了。

nat 方式：

这种方式下 host 内部出现了一个虚拟的网卡 vmnet8（默认情况下），如果你有过 做 nat 服务器的经验，这里的 vmnet8 就相当于连接到内网的网卡，而虚拟机本身则相当于运行在内网上的机器，虚拟机内的网卡（eth0）则独立于 vmnet8。

你会发现在这种方式下，vmware 自带的 dhcp 会默认地加载到 vmnet8 界面上，这样 虚拟机就可以使用 dhcp 服务。更为重要的是，vmware 自带了 nat 服务，提供了从 vmnet8 到外网的地址转换，所以这种情况是一个实实在在的

nat 服务器在运行，只不过是供虚拟机用的。

很显然，如果你只有一个外网地址，此种方式很合适。

hostonly:

这应该是最为灵活的方式，有兴趣的话可以进行各种网络实验。和 **nat** 唯一的不同的是，此 种方式下，没有地址转换服务，因此，模认情况下，虚拟机只能到主机访问，这也是 **hostonly** 的名字的意义。

默认情况下，也会有一个 **dhcp** 服务加载到 **vmnet1** 上。这样连接到 **vmnet1** 上的虚拟机仍然可以设置成 **dhcp**，方便系统的配置。

是不是这种方式就没有办法连接到外网呢，当然不是，事实上，这种方式更为灵活，你可以使用自己的方式，从而达到最理想的配置，例如：

a. 使用自己 **dhcp** 服务：首先停掉 **vmware** 自带的 **dhcp** 服务，使 **dhcp** 服务更为统一。

b. 使用自己的 **nat**，方便加入防火墙。**windows host** 可以做 **nat** 的方法很多，简单的如 **windows xp** 的 **internet** 共享，复杂的如 **windows server** 里的 **nat** 服务。

c. 使用自己的防火墙。因为你可以完全控制 **vmnet1**，你可以加入（或试验）防火墙在 **vmnet1** 和外网的网卡间。

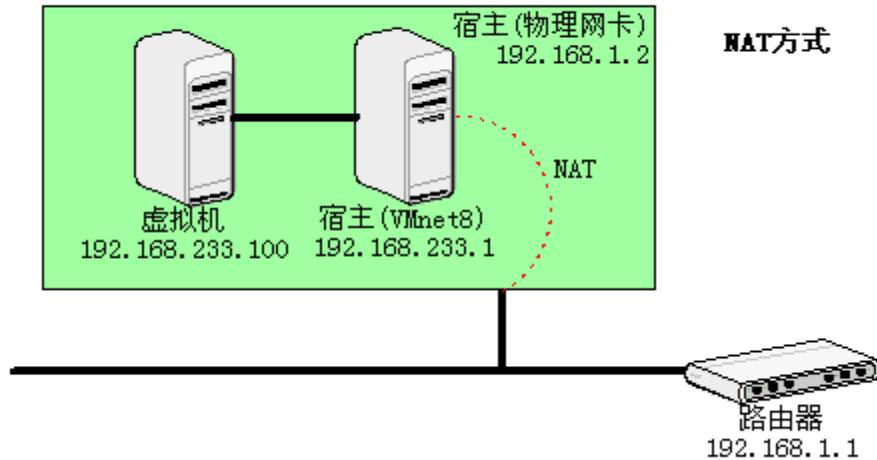
从以上可以看出，**hostonly** 这种模式和普通的 **nat server** 带整个内网上网的情形类似，因此你可以方便的进行与之有关的实验，比如防火墙的设置等。

3.第三套说法

VMware 支持三种类型的网络：**NAT**, **Bridged**, **Host-only**。

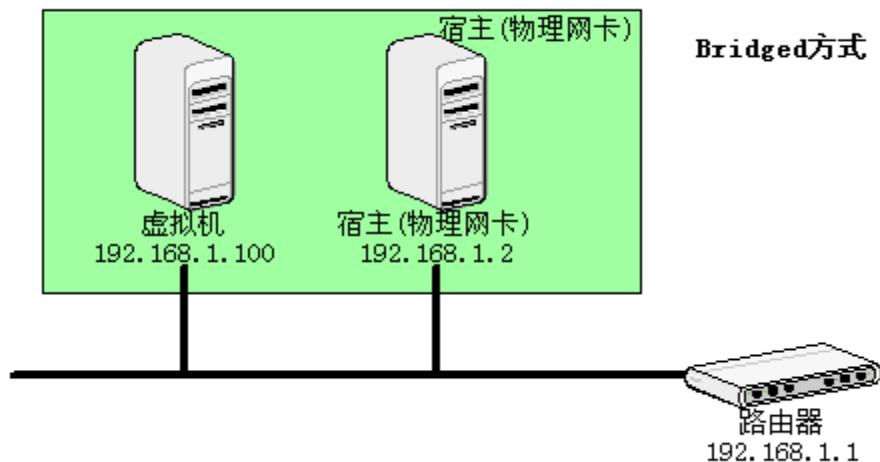
NAT

这种方式下，虚拟机的网卡连接到宿主的 **VMnet8** 上。此时系统的 **VMWare NAT Service** 服务就充当了路由器的作用，负责将虚拟机发到 **VMnet8** 的包进行地址转换之后发到实际的网络上，再将实际网络上返回的包进行地址转换后通过 **VMnet8** 发送给虚拟机。**VMWare DHCP Service** 负责为虚拟机提供 **DHCP** 服务。



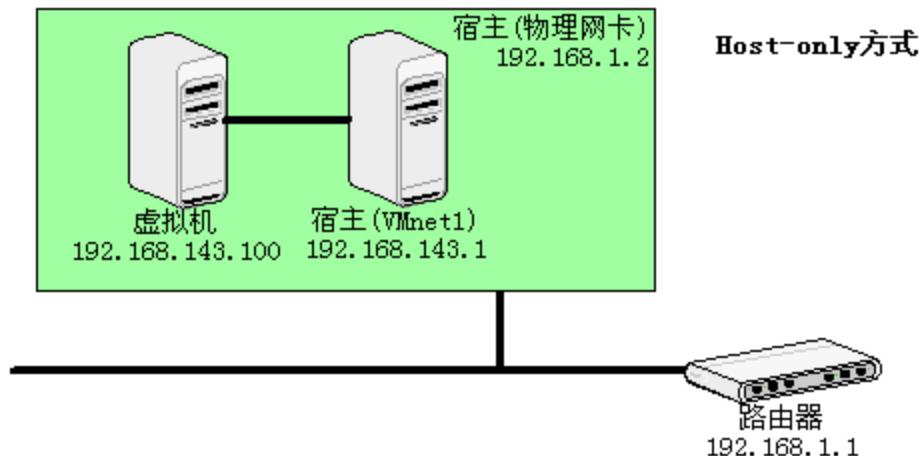
Bridged

这种方式下，虚拟机就像一台真正的计算机一样，直接连接到实际的网络上，与宿主机没有任何联系。



Host-only

这种方式下，虚拟机的网卡连接到宿主的 VMnet1 上，但系统并不为虚拟机提供任何路由服务，因此虚拟机只能和宿主机进行通信，而不能连接到实际网络上。



2) replicate physical network connection state 选项解释：

Select if the virtual machine uses a bridged network connection and if you use the virtual machine on a laptop or other mobile device. As you move from one wired or wireless network to another, the IP address is automatically renewed.
单机多网卡 或者 笔记本和移动设备使用虚拟机时，当在有线网络和无线网络切换时，勾选了 Replicate physical network connection state，虚拟机网卡的 ip 地址会自动更新，不需要重新设置。

转：<http://zhzhzh-43.blog.163.com/blog/static/126697371200910259452260/>
<http://blog.csdn.net/kofterry/article/details/5184789>

三十四、附录 Python 学习资源大全

强烈推荐-**易桌寒**大侠的学习资源整理

本文主要针对零基础想学习Python者。本文由@史江歌（shijiangge@gmail.com）根据互联网资料收集整理而成，感谢互联网，感谢各位的分享。鸣谢！本文会不断更新。

1) 一、学会使用搜索引擎，尤其是 Google。推荐阅读文章：

[《如何用好 Google 搜索引擎？》](#) --知乎问答@崔凯

[《十大高明的 Google 搜索技巧》](#) by-雪影蓝风

《如何用好 Google》 by-李笑来

2) 二、学会提问

推荐阅读《提问的智慧》 --by 啄木鸟社区

3) 三、书籍推荐：

1. Python 官方文档 [点击打开链接](#) 英语不好的，暂时就先别看了，以后一定肯定要加强英语学习哦;)。
2. 对于零基础者，推荐《How to think like a computer scientist》[点击打开链接](#)，对应中文版本《Python 学习笔记--皮大庆》，非常适合零基础入门。
3. 对于零基础者，推荐《Python 编程实践》[点击打开链接](#)
4. 对于零基础者，推荐《简明 Python 教程》[点击打开链接](#)
5. 对于零基础者，推荐《笨办法学 Python (Learn Python The Hard Way)》[点击打开链接](#)
6. 对于有其他编程语言基础者，推荐《Python 基础教程》[点击打开链接](#)
7. 更多书籍推荐，知乎问答：关于 Python 的经典入门书籍有哪些？[点击打开链接](#)
8. 更多书籍推荐，V2EX 问答：文科生想学 Python [点击打开链接](#)

4) 四、视频推荐：

麻省理工公开课《计算机科学及编程导论》，该课程针对零基础者 [点击打开链接](#) 网易公开课。

课程介绍 这门课程适用于那些拥有很少或没有编程经验的学生,它致力于使学生理解计算机在解决问题中的作用，并且帮助学生，不论其专业，使他们对于能够完成有用的小程序的目标充满信心。

5) 五、学习方法

1. 知乎问答：零基础，应当如何开始学习 Python ？[点击打开链接](#)--by@黎敏

虽然我不是 Python 高手，但我是零基础，之前会的都是软件 PS, PPT 之类。

如果目的是想成为程序员，参考教学大纲。

如果只是学程序，理解科技，解决工作问题，我的方式可以参考使用：

1. 找到合适的入门书籍，大致读一次，循环啊判断啊，常用类啊，搞懂（太难的跳过）
2. 做些简单习题，字符串比较，读取日期之类 《Python Cookbook》不错（太难太无趣的，再次跳过，保持兴趣是最重要的，不会的以后可以再学）

3. 加入 Python 讨论群，态度友好笑眯眯（很重要，这样高手才会耐心纠正你错误常识）。很多小问题，纠结许久，对方一句话点播思路，真的节约你很多时间。耐心指教我的好人，超级超级多谢。
4. 解决自己电脑问题。比如下载美剧，零散下载了 2, 4, 5, 8 集，而美剧共 12 集，怎样找出漏下的那几集？然后问题分解，1 读取全部下载文件名，2 提取集的数字，3 数字排序和 (1-12) 对比，找出漏下的。
5. 时刻记住目的，不是为了当程序员，是为了解决问题。

比如，想偷懒抓网页内容，用 `urllib` 不行，用 `request` 也不行，才发现抓取内容涉及那么多方面 (`cookie`, `header`, `SSL`, `url`, `javascript` 等等)，当然可以听人家劝，回去好好读书，从头读。

或者，不求效率，只求解决，用 `ie` 打开网页再另存为行不行？`ie` 已经渲染过全部结果了。

问题变成：1--打开指定的 10 个网页（一行代码就行）。更复杂的想保存呢？利用已经存在的包，比如 `PAM30`（我用的是 `Python3`），直接打开 `ie`，用函数 `outHTML` 另存为文本，再用搜索函数（`str` 搜索也行，`re` 正则也行）找到数据。简单吧？而且代码超级短。

6. 保持兴趣，用最简单的方式解决问题，什么底层驱动，各种交换，留给大牛去写吧。我们利用已经有的包完成。
7. 耐心读文档，并且练习快速读文档。拿到新包，找到自己所需要的函数，是需要快速读一次的。这个不难，读函数名，大概能猜到是干嘛的，然后看看返回值，能判断是不是自己需要的。
8. 写帮助文件和学习笔记，并发布共享。教别人的时候，其实你已经自己再次思考一次了。

我觉得学程序就像学英文，把高频率的词（循环，判断，常用包，常用函数）搞懂，就能拼装成自己想要的软件。

然后，<http://stackoverflow.com> 和 <http://Google.com> 是很好用的。

然后，坚持下去。

一定要保持兴趣，太复杂的跳过，就像小学数学，小学英语，都是由简入深。

网络很平面，无数国际大牛著作好书，关于 `Python`，算法，电脑，网络，或者程序员思路，或者商业思维（浪潮之巅是本好书）等等，还有国际名校的网络公开课（中英文字幕翻译完毕，观看不是难事），讲计算机，网络，安全，或者安卓系统，什么都有，只要能持续保持兴趣，一点点学习下去，不是难事。

所有天才程序员，都曾是儿童，回到儿童思维来理解和学习。觉得什么有趣，先学，不懂的，先放着，遇到问题再来学，效果更好。

唯一建议是，不要太贪心，耐心学好一门优雅的语言，再学其它。虽然 `Javascript` 做特效很炫，或提某问题时，有大牛建议，用 `Ruby` 来写更好之类，不要改方向。就像老笑话：“要学习递归，必须首先理解递归。”然后死循环一直下去。坚持学好一门语言，再研究其他。

即使一门语言，跟网络，数据库等等相关的部分，若都能学好，再学其他语言，是很快的事情。

另外就是，用学英文的耐心来学计算机，英文遇到不懂的词，抄下，查询。

`python` 里，看到 `Http`，查查定义，看到 `outHtml`，查查定义，跟初学英语时候一样，不要直接猜意思，因为精确描述性定义，跟含糊自然语有区别的。而新人瞎猜，很容易错误理解，`wiki`, `google` 很有用。

2. 芝麻问答：如何学习 Python [点击打开链接](#)作者：@halida

如何学习 python

我们假设你是一个初级程序员，只懂得一点点的基础知识，希望能够用 python 来做开发。这篇文档就是为了满足以上目标而写的。

大纲

按照这篇文档所指示的任务过一遍，你就能够做到：

熟悉 python 语言，以及学会 python 的编码方式。熟悉 python 库，遇到开发任务的时候知道如何去找对应的模块。知道如何查找和获取第三方的 python 库，以应付开发任务。学习步骤

安装开发环境

如果你在 window 下，去下载 python 安装起来，然后运行 python，进入 python 解释环境。

如果你在 ubuntu 下，执行: `sudo apt-get install python`，然后在命令行下运行 python，进入 python 解释环境。

学习方法

作为一名成熟的开发人员，我学习新东西(假设是 PyQt)的习惯方式是：

直接用 google 搜索 PyQt 的官方网站。按照官方网站的说明，下载 PyQt。(如果是用 ubuntu，看看软件库里面是否有足够新的版本) 下载过程中，开始阅读官方网站上面的教程。一边看教程，一遍按照教程使用 PyQt。如果发现教程不够全面，用 google 搜索是否有对应的教学书籍可以看。示例学习完毕，开发一个玩具程序，用来检验自己是否需要用到的功能。开始学习 python

我建议你学习的过程也按照上面来，首先过一遍 python 官方文档：

<http://docs.python.org/tutorial/index.html>

然后做 <http://www.pythonchallenge.com/> 这个网站上面的题目。

如果卡在某一关太久，可以看答案(google python challenge answer)，做完后看看别人的编码方式和自己有什么区别。

小项目

做完一遍后，你会发现已经熟悉了基本的 python 开发。然后做点小项目吧。这里是一些题目，挑感兴趣的去做。

- 写一个简单的计算器/记账软件/扫雷游戏(用 PyQt 库做界面)
- 写一个聊天室网站(用 webpy 框架, jquery 刷新新的回复)
- 写一个爬虫，获取豆瓣上面所有用户的地点，画地点分布的直方图(用 lxml 解析，保存数据到 sqlite 里面去，用 matplotlib 画图)

FAQ

Q: 遇到了问题，到哪里求助？

A: 上 <http://groups.google.com/group/python-cn> (需要翻墙) 或者 <http://stackoverflow.com> 提问就可以了。

Q: 如何查找 python 的某个功能？

A: 看官方文档. <http://docs.python.org/library/index.html>

Q: 如何用 python 完成一个任务(比如写网站)？

A: google: python 写网站, 或者 google: python web development.

3. 知乎问答：你是如何自学 Python 的？[点击打开链接](#) 作者:@成增存

阶段一： 前辈领进门。 第一眼是一个前辈给我看了看 Python 的代码，因为自己最早学习的是 Java 语言，第一眼看到 Python 的时候，前辈说，“Python 是面向对象的”，然后就 打印了几句代码。可我怎么也看不出到底哪里是“面向对象”的。前辈说“在 Python 里，一切皆对象”，我才有点领悟的说道：“原来把什么看成对象，就是面向对象”，哈哈。

阶段二： 开始看的是《Dive into Python》、《Leaning python》、《Python Doc》。 因为有很多语言学习经历，很快就完成了初步语法和使用学习。太复杂的特性还是在使用中逐步掌握的。

阶段三： 开始使用 Python 做自己的一些日常工作。 比如 Python 搜索文件，Python 批量处理等，使用最多的还是 re 模块和 socket 相关模块。写了大量的例子，让自己对 Python 更加喜爱，也更加熟悉。此时翻阅最多的是《Python Doc》的指南。不管是语言参考、库参考、Demo 参考，都有大量可使用的内容，内容质量很高堪比 JDK。

阶段四： 生产上马。 开始使用 Django, Flask, Tornado 开发一些 web 应用，写一些日常使用的工具包等。逐步提升设计能力，和整体代码的管理能力。

阶段五： 更合理的分配好 C, Java, Python 三门语言各自擅长的部分。 把合适的语言用到合适的地方。尽管一门语言有时候可以搞定所有的，但用擅长的语言解决合适的问题才是效率最高的。这也是“Python 的大道至简”的理念带给我的帮助和认识。

PS: 其中过程中主要的一些 方法：

1. 看书。学习的基础。
2. 自己本地练习。编程还是要实践出真知。
3. 资料查询。google, stackoverflow 等多关注。
4. 交流。各种论坛上的 python group, 论坛。最早去的 CU, JavaEye, 不过现在貌似去的少了。**google group** 必须订阅。

学习+实践+总结，掌握语言的法宝。

6) 六、Just Do It

看了这么多资料了，还没动手敲代码？

好吧，最好一篇文章《我在学习编程中犯的两个 最大错误 》[点击打开链接](#)

Just do it 。动手敲代码吧。

注：本文由@史江歌（shijiangge@gmail.com）根据互联网资料收集整理而成。

7) 七、更新

Python 技术博客、招聘、开源软件、Python 中文网站导航 [点击打开链接](#)

遇到 python 问题怎么样解决？python help dir stackoverflow docs google [点击打开链接](#)

楚广明新员工培训专用（借用了无数开源先驱的文档与资源，有些出处以无法考证。向开源文档的先驱致敬！）

在线视频教学(英文) [点击打开 codecademy](#)

英文 [pythonforbeginners](#)

英文 How do I learn Python from zero to web development? (各种电子书推荐) [点击打开链接](#)