

Python 3教程



廖雪峰

目录

前言	0
简介	1
安装	2
解释器	2.1
第一个程序	3
使用文本编辑器	3.1
代码运行助手	3.2
输入和输出	3.3
基础	4
数据类型和变量	4.1
字符串和编码	4.2
使用list和tuple	4.3
条件判断	4.4
循环	4.5
使用dict和set	4.6
函数	5
调用函数	5.1
定义函数	5.2
函数的参数	5.3
递归函数	5.4
高级特性	6
切片	6.1
迭代	6.2
列表生成式	6.3
生成器	6.4
迭代器	6.5
函数式编程	7
模块	8
面向对象编程	9
面向对象高级编程	10
错误、调试和测试	11
IO编程	12
正则表达式	13
常用内建模块	14
常用第三方模块	15
virtualenv	16
图形界面	17

网络编程	18
电子邮件	19
访问数据库	20
Web开发	21
异步IO	22
实战	23
FAQ	24
期末总结	25

前言

这是小白的Python新手教程，具有如下特点：

中文，免费，零起点，完整示例，基于最新的Python 3版本。

Python是一种计算机程序设计语言。你可能已经听说过很多种流行的编程语言，比如非常难学的C语言，非常流行的Java语言，适合初学者的Basic语言，适合网页编程的JavaScript语言等等。

那Python是一种什么语言？

首先，我们普及一下编程语言的基础知识。用任何编程语言来开发程序，都是为了让计算机干活，比如下载一个MP3，编写一个文档等等，而计算机干活的CPU只认识机器指令，所以，尽管不同的编程语言差异极大，最后都得“翻译”成CPU可以执行的机器指令。而不同的编程语言，干同一个活，编写的代码量，差距也很大。

比如，完成同一个任务，C语言要写1000行代码，Java只需要写100行，而Python可能只要20行。

所以Python是一种相当高级的语言。

你也许会问，代码少好不好？代码少的代价是运行速度慢，C程序运行1秒钟，Java程序可能需要2秒，而Python程序可能就需要10秒。

那是不是越低级的程序越难学，越高级的程序越简单？表面上来说，是的，但是，在非常高的抽象计算中，高级的Python程序设计也是非常难学的，所以，高级程序语言不等于简单。

但是，对于初学者和完成普通任务，Python语言是非常简单易用的。连Google都在大规模使用Python，你就不用担心学了会没用。

用Python可以做什么？可以做日常任务，比如自动备份你的MP3；可以做网站，很多著名的网站包括YouTube就是Python写的；可以做网络游戏的后台，很多在线游戏的后台都是Python开发的。总之就是能干很多很多事啦。

Python当然也有不能干的事情，比如写操作系统，这个只能用C语言写；写手机应用，只能用Swift/Objective-C（针对iPhone）和Java（针对Android）；写3D游戏，最好用C或C++。

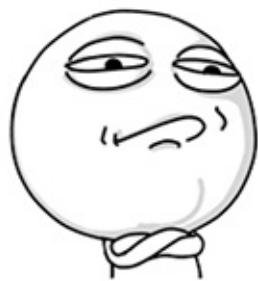
如果你是小白用户，满足以下条件：

- 会使用电脑，但从来没写过程序；
- 还记得初中数学学的方程式和一点点代数知识；
- 想从编程小白变成专业的软件架构师；
- 每天能抽出半个小时学习。

不要再犹豫了，这个教程就是为你准备的！

准备好了吗？

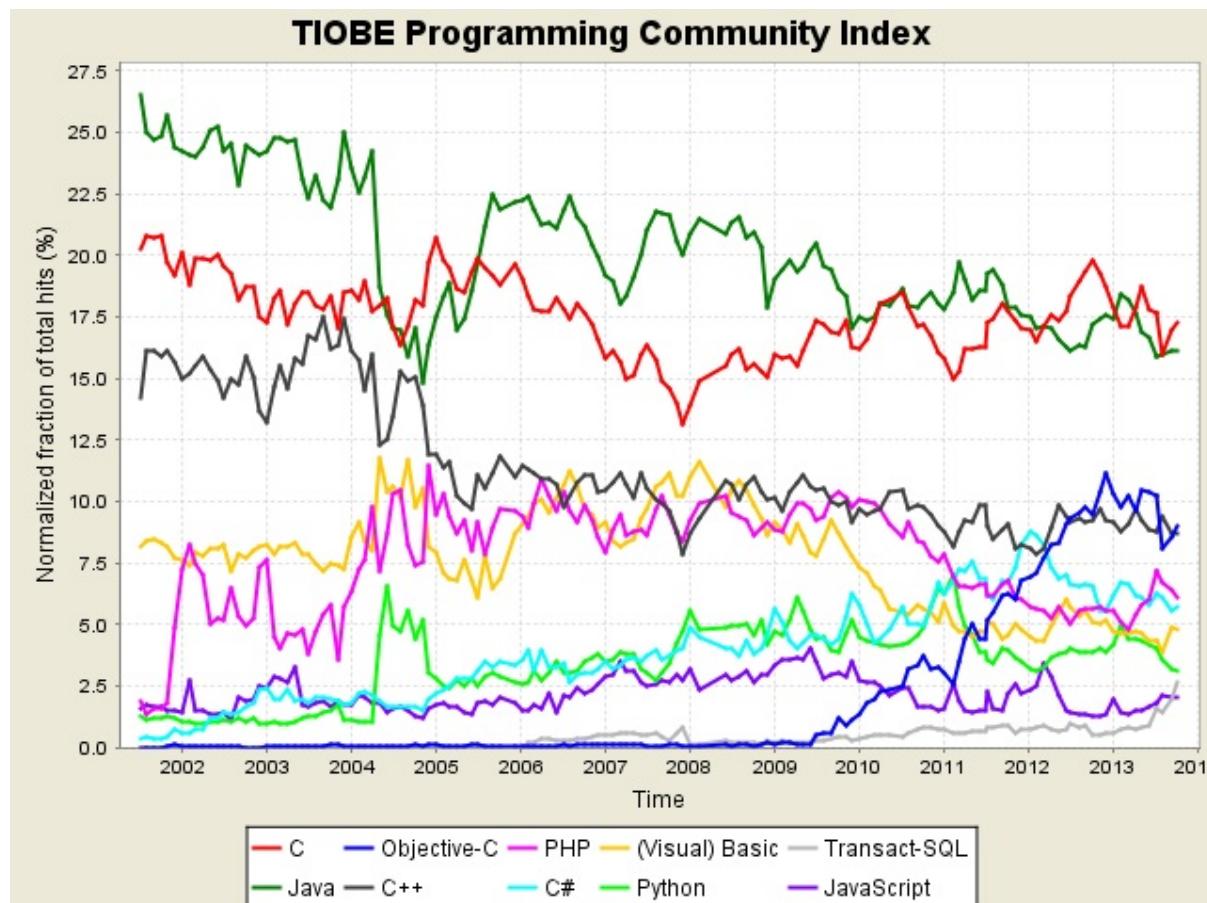
CHALLENGE ACCEPTED !



简介

Python是著名的“龟叔”Guido van Rossum在1989年圣诞节期间，为了打发无聊的圣诞节而编写的一个编程语言。

现在，全世界差不多有600多种编程语言，但流行的编程语言也就那么20来种。如果你听说过TIOBE排行榜，你就能知道编程语言的大致流行程度。这是最近10年最常用的10种编程语言的变化图：



总的来说，这几种编程语言各有千秋。C语言是可以用来编写操作系统的贴近硬件的语言，所以，C语言适合开发那些追求运行速度、充分发挥硬件性能的程序。而Python是用来编写应用程序的高级编程语言。

当你用一种语言开始作真正的软件开发时，你除了编写代码外，还需要很多基本的已经写好的现成的东西，来帮助你加快开发进度。比如说，要编写一个电子邮件客户端，如果先从最底层开始编写网络协议相关的代码，那估计一年半载也开发不出来。高级编程语言通常都会提供一个比较完善的基础代码库，让你能直接调用，比如，针对电子邮件协议的SMTP库，针对桌面环境的GUI库，在这些已有的代码库的基础上开发，一个电子邮件客户端几天就能开发出来。

Python就为我们提供了非常完善的基础代码库，覆盖了网络、文件、GUI、数据库、文本等大量内容，被形象地称作“内置电池（batteries included）”。用Python开发，许多功能不必从零编写，直接使用现成的即可。

除了内置的库外，Python还有大量的第三方库，也就是别人开发的，供你直接使用的东西。当然，如果你开发的代码通过很好的封装，也可以作为第三方库给别人使用。

许多大型网站就是用Python开发的，例如YouTube、Instagram，还有国内的豆瓣。很多大公司，包括Google、Yahoo等，甚至NASA（美国航空航天局）都大量地使用Python。

龟叔给Python的定位是“优雅”、“明确”、“简单”，所以Python程序看上去总是简单易懂，初学者学Python，不但入门容易，而且将来深入下去，可以编写那些非常非常复杂的程序。

总的来说，Python的哲学就是简单优雅，尽量写容易看明白的代码，尽量写少的代码。如果一个资深程序员向你炫耀他写的晦涩难懂、动不动就几万行的代码，你可以尽情地嘲笑他。

那Python适合开发哪些类型的应用呢？

首选是网络应用，包括网站、后台服务等等；

其次是许多日常需要的小工具，包括系统管理员需要的脚本任务等等；

另外就是把其他语言开发的程序再包装起来，方便使用。

最后说说Python的缺点。

任何编程语言都有缺点，Python也不例外。优点说过了，那Python有哪些缺点呢？

第一个缺点就是运行速度慢，和C程序相比非常慢，因为Python是解释型语言，你的代码在执行时会一行一行地翻译成CPU能理解的机器码，这个翻译过程非常耗时，所以很慢。而C程序是运行前直接编译成CPU能执行的机器码，所以非常快。

但是大量的应用程序不需要这么快的运行速度，因为用户根本感觉不出来。例如开发一个下载MP3的网络应用程序，C程序的运行时间需要0.001秒，而Python程序的运行时间需要0.1秒，慢了100倍，但由于网络更慢，需要等待1秒，你想，用户能感觉到1.001秒和1.1秒的区别吗？这就好比F1赛车和普通的出租车在北京三环路上行驶的道理一样，虽然F1赛车理论时速高达400公里，但由于三环路堵车的时速只有20公里，因此，作为乘客，你感觉的时速永远是20公里。



第二个缺点就是代码不能加密。如果要发布你的Python程序，实际上就是发布源代码，这一点跟C语言不同，C语言不用发布源代码，只需要把编译后的机器码（也就是你在Windows上常见的xxx.exe文件）发布出去。要从机器码反推出C代码是不可能的，所以，凡是编译型的语言，都没有这个问题，而解释型的语言，则必须把源码发布出去。

这个缺点仅限于你要编写的软件需要卖给别人挣钱的时候。好消息是目前的互联网时代，靠卖软件授权的商业模式越来越少了，靠网站和移动应用卖服务的模式越来越多了，后一种模式不需要把源码给别人。

再说了，现在如火如荼的开源运动和互联网自由开放的精神是一致的，互联网上有无数非常优秀的像Linux一样的开源代码，我们千万不要高估自己写的代码真的有非常大的“商业价值”。那些大公司的代码不愿意开放的更重要的原因是代码写得太烂了，一旦开源，就没人敢用他们的产品了。



当然，Python还有其他若干小缺点，请自行忽略，就不一一列举了。

安装

因为Python是跨平台的，它可以运行在Windows、Mac和各种Linux/Unix系统上。在Windows上写Python程序，放到Linux上也是能够运行的。

要开始学习Python编程，首先就得把Python安装到你的电脑里。安装后，你会得到Python解释器（就是负责运行Python程序的），一个命令行交互环境，还有一个简单的集成开发环境。

安装Python 3.6

目前，Python有两个版本，一个是2.x版，一个是3.x版，这两个版本是不兼容的。由于3.x版越来越普及，我们的教程将以最新的Python 3.6版本为基础。请确保你的电脑上安装的Python版本是最新的3.6.x，这样，你才能无痛学习这个教程。

在Mac上安装Python

如果你正在使用Mac，系统是OS X 10.8~10.10，那么系统自带的Python版本是2.7。要安装最新的Python 3.6，有两个方法：

方法一：从Python官网下载Python 3.6的[安装程序](#)（网速慢的同学请移步[国内镜像](#)），双击运行并安装；

方法二：如果安装了Homebrew，直接通过命令 `brew install python3` 安装即可。

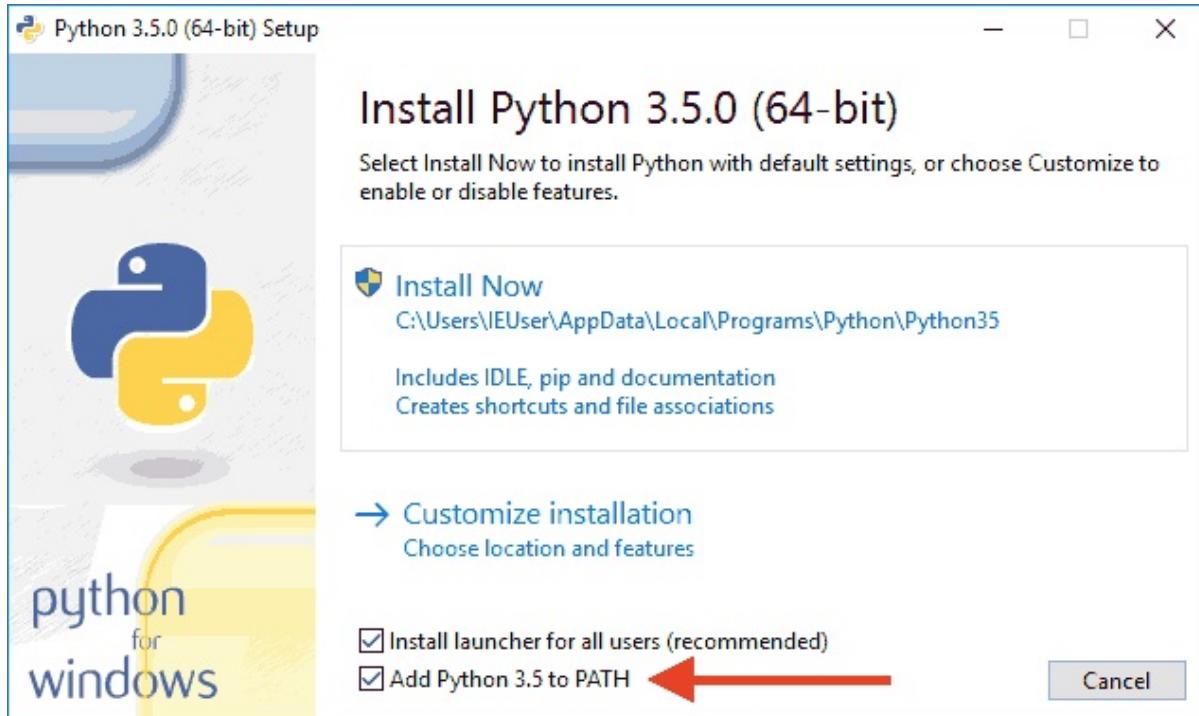
在Linux上安装Python

如果你正在使用Linux，那我可以假定你有Linux系统管理经验，自行安装Python 3应该没有问题，否则，请换回Windows系统。

对于大量的目前仍在使用Windows的同学，如果短期内没有打算换Mac，就可以继续阅读以下内容。

在Windows上安装Python

首先，根据你的Windows版本（64位还是32位）从Python的官方网站下载Python 3.6对应的[64位安装程序](#)或[32位安装程序](#)（网速慢的同学请移步[国内镜像](#)），然后，运行下载的EXE安装包：



特别要注意勾上 Add Python 3.6 to PATH，然后点“Install Now”即可完成安装。

运行Python

安装成功后，打开命令提示符窗口，敲入python后，会出现两种情况：

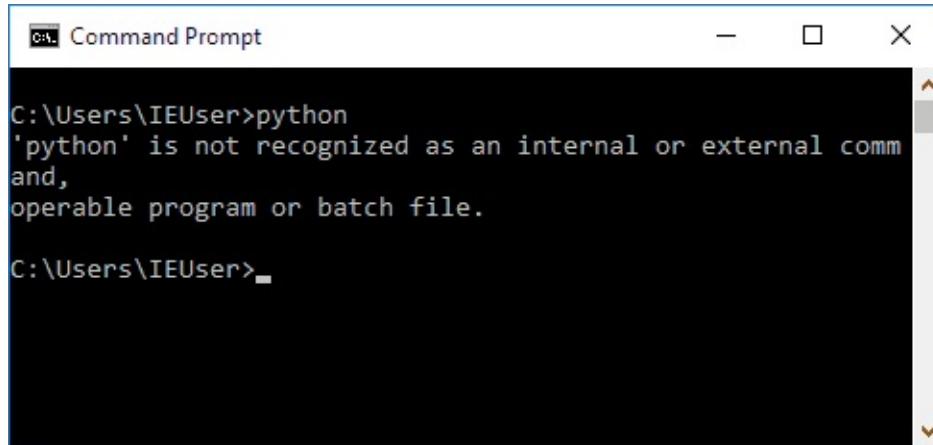
情况一：

看到上面的画面，就说明Python安装成功！

你看到提示符 >>> 就表示我们已经在Python交互式环境中了，可以输入任何Python代码，回车后会立刻得到执行结果。现在，输入 exit() 并回车，就可以退出Python交互式环境（直接关掉命令行窗口也可以）。

情况二：得到一个错误：

```
'python' 不是内部或外部命令，也不是可运行的程序或批处理文件。
```



这是因为Windows会根据一个 Path 的环境变量设定的路径去查找 python.exe , 如果没找到, 就会报错。如果在安装时漏掉了勾选 Add Python 3.6 to PATH , 那就要手动把 python.exe 所在的路径添加到Path中。

如果你不知道怎么修改环境变量, 建议把Python安装程序重新运行一遍, 务必记得勾上 Add Python 3.6 to PATH 。

视频演示:

小结

学会如何把Python安装到计算机中, 并且熟练打开和退出Python交互式环境。

在Windows上运行Python时, 请先启动命令行, 然后运行 python 。

在Mac和Linux上运行Python时, 请打开终端, 然后运行 python3 。

解释器

当我们编写Python代码时，我们得到的是一个包含Python代码的以`.py`为扩展名的文本文件。要运行代码，就需要Python解释器去执行`.py`文件。

由于整个Python语言从规范到解释器都是开源的，所以理论上，只要水平够高，任何人都可以编写Python解释器来执行Python代码（当然难度很大）。事实上，确实存在多种Python解释器。

C Python

当我们从[Python官方网站](#)下载并安装好Python 3.x后，我们就直接获得了一个官方版本的解释器：CPython。这个解释器是用C语言开发的，所以叫CPython。在命令行下运行`python`就是启动CPython解释器。

CPython是使用最广的Python解释器。教程的所有代码也都在CPython下执行。

I Python

I Python是基于CPython之上一个交互式解释器，也就是说，I Python只是在交互方式上有所增强，但是执行Python代码的功能和CPython是完全一样的。好比很多国产浏览器虽然外观不同，但内核其实都是调用了IE。

CPython用`>>>`作为提示符，而I Python用`In [序号]:`作为提示符。

PyPy

PyPy是另一个Python解释器，它的目标是执行速度。PyPy采用[JIT技术](#)，对Python代码进行动态编译（注意不是解释），所以可以显著提高Python代码的执行速度。

绝大部分Python代码都可以在PyPy下运行，但是PyPy和CPython有一些是不同的，这就导致相同的Python代码在两种解释器下执行可能会有不同的结果。如果你的代码要放到PyPy下执行，就需要了解[PyPy和CPython的不同点](#)。

Jython

Jython是运行在Java平台上的Python解释器，可以直接把Python代码编译成Java字节码执行。

IronPython

IronPython和Jython类似，只不过IronPython是运行在微软.NET平台上的Python解释器，可以直接把Python代码编译成.NET的字节码。

小结

Python的解释器很多，但使用最广泛的还是CPython。如果要和Java或.NET平台交互，最好的办法不是用Jython或IronPython，而是通过网络调用来交互，确保各程序之间的独立性。

本教程的所有代码只确保在CPython 3.x版本下运行。请务必在本地安装CPython（也就是从Python官方网站下载的安装程序）。

第一个程序

在正式编写第一个Python程序前，我们先复习一下什么是命令行模式和Python交互模式。

命令行模式

在Windows开始菜单选择“命令提示符”，就进入到命令行模式，它的提示符类似 c:\>：

```
|Command Prompt - □ X
|
|Microsoft Windows [Version 10.0.0]
|(c) 2015 Microsoft Corporation. All rights reserved.
|
|C:\> _
```

Python交互模式

在命令行模式下敲命令 python，就看到类似如下的一堆文本输出，然后就进入到Python交互模式，它的提示符是 >>>。

```
|Command Prompt - python - □ X
|
|Microsoft Windows [Version 10.0.0]
|(c) 2015 Microsoft Corporation. All rights reserved.
|
|C:\> python
|Python 3.6 ... on win32
|Type "help", ... for more information.
|>>> _
```

在Python交互模式下输入 exit() 并回车，就退出了Python交互模式，并回到命令行模式：

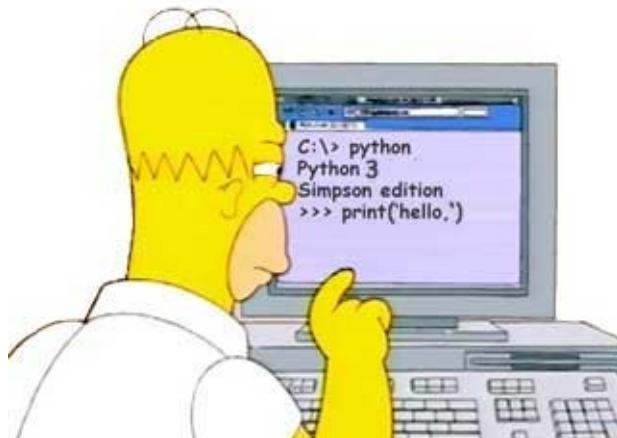
```
|Command Prompt - □ X
|
|Microsoft Windows [Version 10.0.0]
|(c) 2015 Microsoft Corporation. All rights reserved.
|
|C:\> python
|Python 3.6 ... on win32
|Type "help", ... for more information.
|>>> exit()
|
|C:\> _
```



也可以直接通过开始菜单选择 Python (command line) 菜单项，直接进入 Python 交互模式，但是输入 `exit()` 后窗口会直接关闭，不会回到命令行模式。

了解了如何启动和退出 Python 的交互模式，我们就可以正式开始编写 Python 代码了。

在写代码之前，请千万不要用“复制”-“粘贴”把代码从页面粘贴到你自己的电脑上。写程序也讲究一个感觉，你需要一个字母一个字母地把代码自己敲进去，在敲代码的过程中，初学者经常会敲错代码：拼写不对，大小写不对，混用中英文标点，混用空格和 Tab 键，所以，你需要仔细地检查、对照，才能以最快的速度掌握如何写程序。



在交互模式的提示符 `>>>` 下，直接输入代码，按回车，就可以立刻得到代码执行结果。现在，试试输入 `100+200`，看看计算结果是不是 300：

```
>>> 100+200
300
```

很简单吧，任何有效的数学计算都可以算出来。

如果要让 Python 打印出指定的文字，可以用 `print()` 函数，然后把希望打印的文字用单引号或者双引号括起来，但不能混用单引号和双引号：

```
>>> print('hello, world')
hello, world
```

这种用单引号或者双引号括起来的文本在程序中叫字符串，今后我们还会经常遇到。

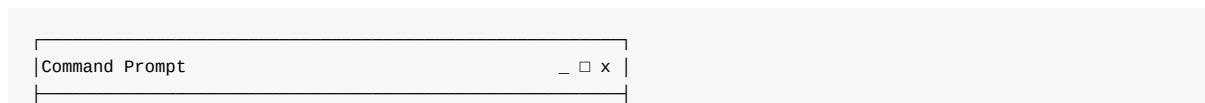
最后，用 `exit()` 退出 Python，我们的第一个 Python 程序完成！唯一的缺憾是没有保存下来，下次运行时还要再输入一遍代码。

命令行模式和 Python 交互模式

请注意区分命令行模式和 Python 交互模式。

在命令行模式下，可以执行 `python` 进入 Python 交互式环境，也可以执行 `python hello.py` 运行一个 `.py` 文件。

执行一个 `.py` 文件只能在命令行模式执行。如果敲一个命令 `python hello.py`，看到如下错误：



```
|Microsoft Windows [Version 10.0.0]
|(c) 2015 Microsoft Corporation. All rights reserved.
|
|C:\> python hello.py
|python: can't open file 'hello.py': [Errno 2] No such
|file or directory
|
|
|
|
|
```

错误提示 `No such file or directory` 说明这个 `hello.py` 在当前目录找不到，必须先把当前目录切换到 `hello.py` 所在的目录下，才能正常执行：

```
[Command Prompt] _ □ x
|
|Microsoft Windows [Version 10.0.0]
|(c) 2015 Microsoft Corporation. All rights reserved.
|
|C:\> cd work
|
|C:\work> python hello.py
>Hello, world!
|
|
|
|
```

此外，在命令行模式运行 `.py` 文件和在Python交互式环境下直接运行Python代码有所不同。Python交互式环境会把每一行Python代码的结果自动打印出来，但是，直接运行Python代码却不会。

例如，在Python交互式环境下，输入：

```
>>> 100 + 200 + 300
600
```

直接可以看到结果 `600`。

但是，写一个 `calc.py` 的文件，内容如下：

```
100 + 200 + 300
```

然后在命令行模式下执行：

```
C:\work>python calc.py
```

发现什么输出都没有。

这是正常的。想要输出结果，必须自己用 `print()` 打印出来。把 `calc.py` 改造一下：

```
print(100 + 200 + 300)
```

再执行，就可以看到结果：

```
C:\work>python calc.py
```

600

最后，Python交互模式的代码是输入一行，执行一行，而命令行模式下直接运行 .py 文件是一次性执行该文件内的所有代码。可见，Python交互模式主要是为了调试Python代码用的，也便于初学者学习，它不是正式运行Python代码的环境！

小结

在Python交互式模式下，可以直接输入代码，然后执行，并立刻得到结果。

在命令行模式下，可以直接运行 .py 文件。

使用文本编辑器

在Python的交互式命令行写程序，好处是一下就能得到结果，坏处是没法保存，下次还想运行的时候，还得再敲一遍。

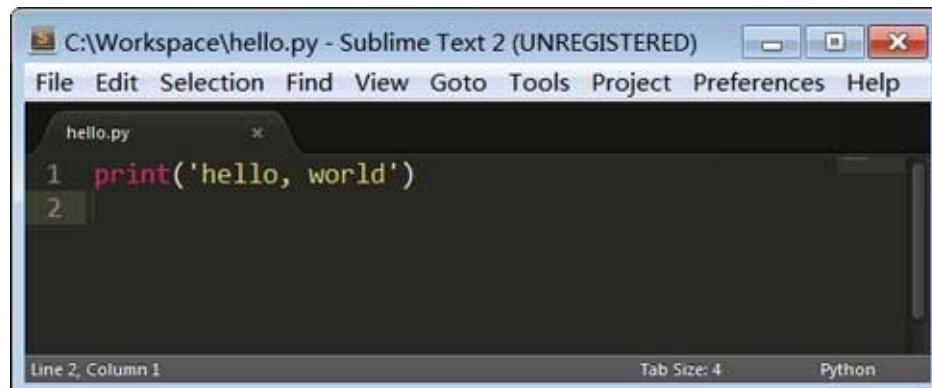
所以，实际开发的时候，我们总是使用一个文本编辑器来写代码，写完了，保存为一个文件，这样，程序就可以反复运行了。

现在，我们就把上次的 'hello, world' 程序用文本编辑器写出来，保存下来。

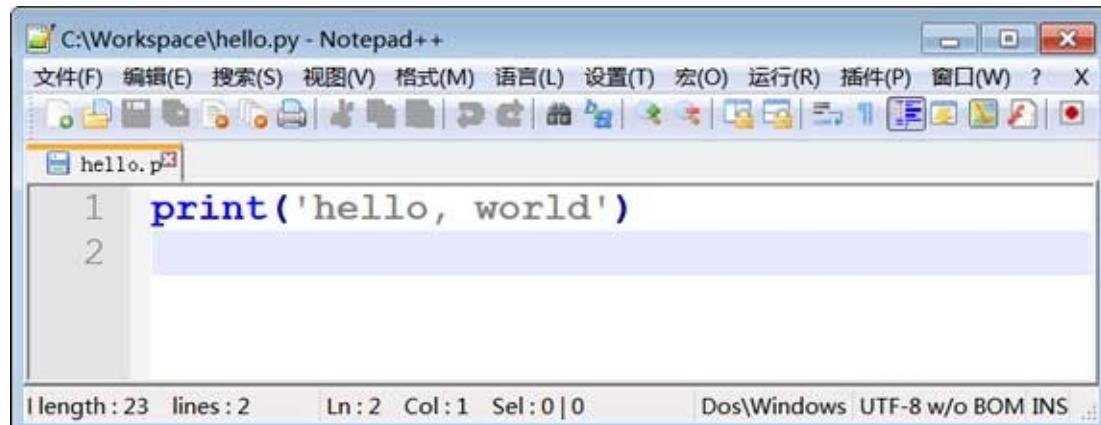
那么问题来了：文本编辑器到底哪家强？

推荐两款文本编辑器：

一个是[Sublime Text](#)，免费使用，但是不付费会弹出提示框：



一个是[Notepad++](#)，免费使用，有中文界面：



请注意，用哪个都行，但是绝对不能用Word和Windows自带的记事本。Word保存的不是纯文本文件，而记事本会自作聪明地在文件开始的地方加上几个特殊字符（UTF-8 BOM），结果会导致程序运行出现莫名其妙的错误。

安装好文本编辑器后，输入以下代码：

```
print('hello, world')
```

注意 `print` 前面不要有任何空格。然后，选择一个目录，例如 `c:\work`，把文件保存为 `hello.py`，就可以打开命令行窗口，把当前目录切换到 `hello.py` 所在目录，就可以运行这个程序了：

```
C:\work>python hello.py
hello, world
```

也可以保存为别的名字，比如 `first.py`，但是必须要以 `.py` 结尾，其他的都不行。此外，文件名只能是英文字母、数字和下划线的组合。

如果当前目录下没有 `hello.py` 这个文件，运行 `python hello.py` 就会报错：

```
C:\Users\IEUser>python hello.py
python: can't open file 'hello.py': [Errno 2] No such file or directory
```

报错的意思就是，无法打开 `hello.py` 这个文件，因为文件不存在。这个时候，就要检查一下当前目录下是否有这个文件了。如果 `hello.py` 存放在另外一个目录下，要首先用 `cd` 命令切换当前目录。

直接运行py文件

有同学问，能不能像`.exe`文件那样直接运行 `.py` 文件呢？在Windows上是不行的，但是，在Mac和Linux上是可以的，方法是在 `.py` 文件的第一行加上一个特殊的注释：

```
#!/usr/bin/env python3

print('hello, world')
```

然后，通过命令给 `hello.py` 以执行权限：

```
$ chmod a+x hello.py
```

就可以直接运行 `hello.py` 了，比如在Mac下运行：

The screenshot shows a Mac OS X terminal window titled "michael — bash — 80x24". The window contains the following text:

```
Last login: Sun Dec 15 18:44:13 on ttys000
Mavericks:- michael$ ./hello.py
hello, world
Mavericks:- michael$
```

小结

用文本编辑器写Python程序，然后保存为后缀为 `.py` 的文件，就可以用Python直接运行这个程序了。

Python的交互模式和直接运行 `.py` 文件有什么区别呢？

直接输入 `python` 进入交互模式，相当于启动了Python解释器，但是等待你一行一行地输入源代码，每输入一行就执行一行。

直接运行 `.py` 文件相当于启动了Python解释器，然后一次性把 `.py` 文件的源代码给执行了，你是没有机会以交互的方式输入源代码的。

用Python开发程序，完全可以一边在文本编辑器里写代码，一边开一个交互式命令窗口，在写代码的过程中，把部分代码粘到命令行去验证，事半功倍！前提是得有个27'的超大显示器！

参考源码

[hello.py](#)

代码运行助手

Python代码运行助手可以让你在线输入Python代码，然后通过本机运行的一个Python脚本来执行代码。原理如下：

- 在网页输入代码：

```
# 测试代码:  
print('Hello, world')
```

 Run

- 点击 Run 按钮，代码被发送到本机正在运行的Python代码运行助手；
- Python代码运行助手将代码保存为临时文件，然后调用Python解释器执行代码；
- 网页显示代码执行结果：

```
# 测试代码:  
print('Hello, world')
```

 Run

Hello, world

下载

点击右键，目标另存为：learning.py

备用下载地址：[learning.py](#)

运行

在存放 learning.py 的目录下运行命令：

```
C:\Users\michael\Downloads> python learning.py
```

如果看到 Ready for Python code on port 39093... 表示运行成功，不要关闭命令行窗口，最小化放到后台运行即可：

```
| Command Prompt  
| Microsoft Windows [Version 10.0.0]  
|(c) 2015 Microsoft Corporation. All rights reserved.
```

```
|  
|C:\Users\michael\Downloads> python learning.py  
|Ready for Python code on port 39093...  
|Press Ctrl + C to exit...  
|  
|  
|  
|  
|
```

试试效果

需要支持HTML5的浏览器：

- IE >= 9
- Firefox
- Chrome
- Sarafi

输入和输出

输出

用 `print()` 在括号中加上字符串，就可以向屏幕上输出指定的文字。比如输出 `'hello, world'`，用代码实现如下：

```
>>> print('hello, world')
```

`print()` 函数也可以接受多个字符串，用逗号“,”隔开，就可以连成一串输出：

```
>>> print('The quick brown fox', 'jumps over', 'the lazy dog')
The quick brown fox jumps over the lazy dog
```

`print()` 会依次打印每个字符串，遇到逗号“,”会输出一个空格，因此，输出的字符串是这样拼起来的：

```
print('The quick brown fox' , 'jumps over' , 'the lazy dog')
          ↓           ↓           ↓           ↓           ↓
The quick brown fox jumps over the lazy dog
```

`print()` 也可以打印整数，或者计算结果：

```
>>> print(300)
300
>>> print(100 + 200)
300
```

因此，我们可以把计算 `100 + 200` 的结果打印得更漂亮一点：

```
>>> print('100 + 200 =', 100 + 200)
100 + 200 = 300
```

注意，对于 `100 + 200`，Python 解释器自动计算出结果 `300`，但是，`'100 + 200 ='` 是字符串而非数学公式，Python 把它视为字符串，请自行解释上述打印结果。

输入

现在，你已经可以用 `print()` 输出你想要的结果了。但是，如果要让用户从电脑输入一些字符怎么办？Python 提供了一个 `input()`，可以让用户输入字符串，并存放到一个变量里。比如输入用户名：

```
>>> name = input()
Michael
```

当你输入 `name = input()` 并按下回车后，Python 交互式命令行就在等待你的输入了。这时，你可以输入任意字符，然后按回车后完成输入。

输入完成后，不会有任何提示，Python 交互式命令行又回到 `>>>` 状态了。那我们刚才输入的内容到哪去了？答案是存放到 `name` 变量里了。可以直接输入 `name` 查看变量内容：

```
>>> name
'Michael'
```

什么是变量？ 请回忆初中数学所学的代数基础知识：

设正方形的边长为 a ，则正方形的面积为 $a \times a$ 。把边长 a 看做一个变量，我们就可以根据 a 的值计算正方形的面积，比如：

若 $a=2$ ，则面积为 $a \times a = 2 \times 2 = 4$ ；

若 $a=3.5$ ，则面积为 $a \times a = 3.5 \times 3.5 = 12.25$ 。

在计算机程序中，变量不仅可以为整数或浮点数，还可以是字符串，因此，`name` 作为一个变量就是一个字符串。

要打印出 `name` 变量的内容，除了直接写 `name` 然后按回车外，还可以用 `print()` 函数：

```
>>> print(name)
Michael
```

有了输入和输出，我们就可以把上次打印 `'hello, world'` 的程序改成有点意义的程序了：

```
name = input()
print('hello,', name)
```

运行上面的程序，第一行代码会让用户输入任意字符作为自己的名字，然后存入 `name` 变量中；第二行代码会根据用户的名字向用户说 `hello`，比如输入 `Michael`：

```
C:\Workspace> python hello.py
Michael
hello, Michael
```

但是程序运行的时候，没有任何提示信息告诉用户：“嘿，赶紧输入你的名字”，这样显得很不友好。幸好，`input()` 可以让你显示一个字符串来提示用户，于是我们把代码改成：

```
name = input('please enter your name: ')
print('hello,', name)
```

再次运行这个程序，你会发现，程序一运行，会首先打印出 `please enter your name:`，这样，用户就可以根据提示，输入名字后，得到 `hello, xxx` 的输出：

```
C:\Workspace> python hello.py
please enter your name: Michael
hello, Michael
```

每次运行该程序，根据用户输入的不同，输出结果也会不同。

在命令行下，输入和输出就是这么简单。

小结

任何计算机程序都是为了执行一个特定的任务，有了输入，用户才能告诉计算机程序所需的信息，有了输出，程序运行后才能告诉用户任务的结果。

输入是Input，输出是Output，因此，我们把输入输出统称为Input/Output，或者简写为IO。

`input()` 和 `print()` 是在命令行下面最基本的输入和输出，但是，用户也可以通过其他更高级的图形界面完成输入和输出，比如，在网页上的一个文本框输入自己的名字，点击“确定”后在网页上看到输出信息。

练习

请利用 `print()` 输出 `1024 * 768 = xxx` :

```
# -*- coding: utf-8 -*-
print('2个数的乘法运算')
A=int(input('乘数A'))
B=int(input('乘数B'))
print('A', 'X', 'B' '=' , A*B)
```

第一个程序

在正式编写第一个Python程序前，我们先复习一下什么是命令行模式和Python交互模式。

命令行模式

在Windows开始菜单选择“命令提示符”，就进入到命令行模式，它的提示符类似 `c:\>`：

```
Command Prompt - □ X
Microsoft Windows [Version 10.0.0]
(c) 2015 Microsoft Corporation. All rights reserved.

C:\> _
```

Python交互模式

在命令行模式下敲命令 `python`，就看到类似如下的一堆文本输出，然后就进入到Python交互模式，它的提示符是 `>>>`。

```
Command Prompt - python - □ X
Microsoft Windows [Version 10.0.0]
(c) 2015 Microsoft Corporation. All rights reserved.

C:\> python
Python 3.6 ... on win32
Type "help", ... for more information.
>>> _
```

在Python交互模式下输入 `exit()` 并回车，就退出了Python交互模式，并回到命令行模式：

```
Command Prompt - □ X
Microsoft Windows [Version 10.0.0]
(c) 2015 Microsoft Corporation. All rights reserved.

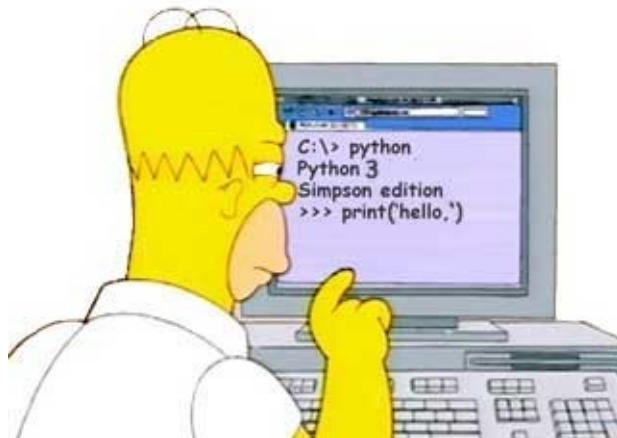
C:\> python
Python 3.6 ... on win32
Type "help", ... for more information.
>>> exit()
C:\> _
```



也可以直接通过开始菜单选择 Python (command line) 菜单项，直接进入 Python 交互模式，但是输入 `exit()` 后窗口会直接关闭，不会回到命令行模式。

了解了如何启动和退出 Python 的交互模式，我们就可以正式开始编写 Python 代码了。

在写代码之前，请千万不要用“复制”-“粘贴”把代码从页面粘贴到你自己的电脑上。写程序也讲究一个感觉，你需要一个字母一个字母地把代码自己敲进去，在敲代码的过程中，初学者经常会敲错代码：拼写不对，大小写不对，混用中英文标点，混用空格和 Tab 键，所以，你需要仔细地检查、对照，才能以最快的速度掌握如何写程序。



在交互模式的提示符 `>>>` 下，直接输入代码，按回车，就可以立刻得到代码执行结果。现在，试试输入 `100+200`，看看计算结果是不是 300：

```
>>> 100+200
300
```

很简单吧，任何有效的数学计算都可以算出来。

如果要让 Python 打印出指定的文字，可以用 `print()` 函数，然后把希望打印的文字用单引号或者双引号括起来，但不能混用单引号和双引号：

```
>>> print('hello, world')
hello, world
```

这种用单引号或者双引号括起来的文本在程序中叫字符串，今后我们还会经常遇到。

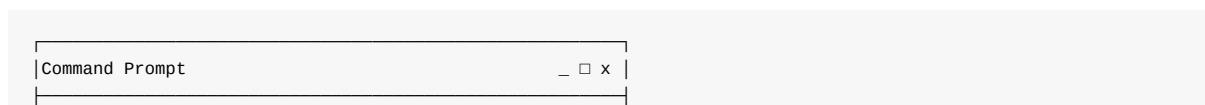
最后，用 `exit()` 退出 Python，我们的第一个 Python 程序完成！唯一的缺憾是没有保存下来，下次运行时还要再输入一遍代码。

命令行模式和 Python 交互模式

请注意区分命令行模式和 Python 交互模式。

在命令行模式下，可以执行 `python` 进入 Python 交互式环境，也可以执行 `python hello.py` 运行一个 `.py` 文件。

执行一个 `.py` 文件只能在命令行模式执行。如果敲一个命令 `python hello.py`，看到如下错误：



```
|Microsoft Windows [Version 10.0.0]
|(c) 2015 Microsoft Corporation. All rights reserved.
|
|C:\> python hello.py
|python: can't open file 'hello.py': [Errno 2] No such
|file or directory
|
|
|
|
|
```

错误提示 `No such file or directory` 说明这个 `hello.py` 在当前目录找不到，必须先把当前目录切换到 `hello.py` 所在的目录下，才能正常执行：

```
[Command Prompt] _ □ x
|
|Microsoft Windows [Version 10.0.0]
|(c) 2015 Microsoft Corporation. All rights reserved.
|
|C:\> cd work
|
|C:\work> python hello.py
>Hello, world!
|
|
|
|
```

此外，在命令行模式运行 `.py` 文件和在Python交互式环境下直接运行Python代码有所不同。Python交互式环境会把每一行Python代码的结果自动打印出来，但是，直接运行Python代码却不会。

例如，在Python交互式环境下，输入：

```
>>> 100 + 200 + 300
600
```

直接可以看到结果 `600`。

但是，写一个 `calc.py` 的文件，内容如下：

```
100 + 200 + 300
```

然后在命令行模式下执行：

```
C:\work>python calc.py
```

发现什么输出都没有。

这是正常的。想要输出结果，必须自己用 `print()` 打印出来。把 `calc.py` 改造一下：

```
print(100 + 200 + 300)
```

再执行，就可以看到结果：

```
C:\work>python calc.py
```

600

最后，Python交互模式的代码是输入一行，执行一行，而命令行模式下直接运行 .py 文件是一次性执行该文件内的所有代码。可见，Python交互模式主要是为了调试Python代码用的，也便于初学者学习，它不是正式运行Python代码的环境！

小结

在Python交互式模式下，可以直接输入代码，然后执行，并立刻得到结果。

在命令行模式下，可以直接运行 .py 文件。

数据类型和变量

数据类型

计算机顾名思义就是可以做数学计算的机器，因此，计算机程序理所当然地可以处理各种数值。但是，计算机能处理的远不止数值，还可以处理文本、图形、音频、视频、网页等各种各样的数据，不同的数据，需要定义不同的数据类型。在Python中，能够直接处理的数据类型有以下几种：

整数

Python可以处理任意大小的整数，当然包括负整数，在程序中的表示方法和数学上的写法一模一样，例如：`1`，`100`，`-8080`，`0`，等等。

计算机由于使用二进制，所以，有时候用十六进制表示整数比较方便，十六进制用`0x`前缀和0-9，`a-f`表示，例如：`0xff00`，`0xa5b4c3d2`，等等。

浮点数

浮点数也就是小数，之所以称为浮点数，是因为按照科学记数法表示时，一个浮点数的小数点位置是可变的，比如，`1.23x109`和`12.3x108`是完全相等的。浮点数可以用数学写法，如`1.23`，`3.14`，`-9.01`，等等。但是对于很大或很小的浮点数，就必须用科学计数法表示，把10用e替代，`1.23x109`就是`1.23e9`，或者`12.3e8`，`0.000012`可以写成`1.2e-5`，等等。

整数和浮点数在计算机内部存储的方式是不同的，整数运算永远是精确的（除法难道也是精确的？是的！），而浮点数运算则可能会有四舍五入的误差。

字符串

字符串是以单引号`'`或双引号`"`括起来的任意文本，比如`'abc'`，`"xyz"`等等。请注意，`'`或`"`本身只是一种表示方式，不是字符串的一部分，因此，字符串`'abc'`只有`a`，`b`，`c`这3个字符。如果`'`本身也是一个字符，那就可以用`'''`括起来，比如`"I'm OK"`包含的字符是`I`，`'`，`m`，空格，`o`，`K`这6个字符。

如果字符串内部既包含`'`又包含`"`怎么办？可以用转义字符`\`来标识，比如：

```
'I\'m \"OK\"!'
```

表示的字符串内容是：

```
I'm "OK"!
```

转义字符`\`可以转义很多字符，比如`\n`表示换行，`\t`表示制表符，字符`\`本身也要转义，所以`\\\`表示的字符就是`\`，可以在Python的交互式命令行用`print()`打印字符串看看：

```
>>> print('I\'m ok.')
I'm ok.
>>> print('I\'m learning\nPython.')
I'm learning
Python.
>>> print('\\\\n\\\\')
\\
\\
```

如果字符串里面有很多字符都需要转义，就需要加很多 `\\`，为了简化，Python还允许用 `r'''` 表示 `'''` 内部的字符串默认不转义，可以自己试试：

```
>>> print('\\\\t\\\\')
\\
>>> print(r'\\\\t\\\\')
\\\\t\\\"
```

如果字符串内部有很多换行，用 `\\n` 写在一行里不好阅读，为了简化，Python允许用 `'''...'''` 的格式表示多行内容，可以自己试试：

```
>>> print('''line1
... line2
... line3''')
line1
line2
line3
```

上面是在交互式命令行内输入，注意在输入多行内容时，提示符由 `>>>` 变为 `...`，提示你可以接着上一行输入，注意 `...` 是提示符，不是代码的一部分：

```
|Command Prompt - python
|_
|>>> print('''line1
|... line2
|... line3''')
|line1
|line2
|line3
|
|>>> _
```

当输入完结束符 `'''` 和括号 `)` 后，执行该语句并打印结果。

如果写成程序并存为 .py 文件，就是：

```
print('''line1
line2
line3''')
```

多行字符串 `'''...'''` 还可以在前面加上 `r` 使用，请自行测试：

```
# -*- coding: utf-8 -*-
print(r'''hello,\nworld'''')
```

布尔值

布尔值和布尔代数的表示完全一致，一个布尔值只有 `True`、`False` 两种值，要么是 `True`，要么是 `False`，在Python中，可以直接用 `True`、`False` 表示布尔值（请注意大小写），也可以通过布尔运算计算出来：

```
>>> True
```

```

True
>>> False
False
>>> 3 > 2
True
>>> 3 > 5
False

```

布尔值可以用 `and`、`or` 和 `not` 运算。

`and` 运算是与运算，只有所有都为 `True`，`and` 运算结果才是 `True`：

```

>>> True and True
True
>>> True and False
False
>>> False and False
False
>>> 5 > 3 and 3 > 1
True

```

`or` 运算是或运算，只要其中有一个为 `True`，`or` 运算结果就是 `True`：

```

>>> True or True
True
>>> True or False
True
>>> False or False
False
>>> 5 > 3 or 1 > 3
True

```

`not` 运算是非运算，它是一个单目运算符，把 `True` 变成 `False`，`False` 变成 `True`：

```

>>> not True
False
>>> not False
True
>>> not 1 > 2
True

```

布尔值经常用在条件判断中，比如：

```

if age >= 18:
    print('adult')
else:
    print('teenager')

```

空值

空值是Python里一个特殊的值，用 `None` 表示。`None` 不能理解为 `0`，因为 `0` 是有意义的，而 `None` 是一个特殊的空值。

此外，Python还提供了列表、字典等多种数据类型，还允许创建自定义数据类型，我们后面会继续讲到。

变量

变量的概念基本上和初中代数的方程变量是一致的，只是在计算机程序中，变量不仅可以是数字，还可以是任意数据类型。

变量在程序中就是用一个变量名表示了，变量名必须是大小写英文、数字和 _ 的组合，且不能用数字开头，比如：

```
a = 1
```

变量 a 是一个整数。

```
t_007 = 'T007'
```

变量 t_007 是一个字符串。

```
Answer = True
```

变量 Answer 是一个布尔值 True。

在Python中，等号 = 是赋值语句，可以把任意数据类型赋值给变量，同一个变量可以反复赋值，而且可以是不同类型的变量，例如：

```
# -*- coding: utf-8 -*-
a = 123 # a是整数
print(a)
a = 'ABC' # a变为字符串
print(a)
```

这种变量本身类型不固定的语言称之为动态语言，与之对应的是静态语言。静态语言在定义变量时必须指定变量类型，如果赋值的时候类型不匹配，就会报错。例如Java是静态语言，赋值语句如下（// 表示注释）：

```
int a = 123; // a是整数类型变量
a = "ABC"; // 错误：不能把字符串赋给整型变量
```

和静态语言相比，动态语言更灵活，就是这个原因。

请不要把赋值语句的等号等同于数学的等号。比如下面的代码：

```
x = 10
x = x + 2
```

如果从数学上理解 $x = x + 2$ 那无论如何是不成立的，在程序中，赋值语句先计算右侧的表达式 $x + 2$ ，得到结果 12，再赋给变量 x。由于 x 之前的值是 10，重新赋值后，x 的值变成 12。

最后，理解变量在计算机内存中的表示也非常 important。当我们写：

```
a = 'ABC'
```

时，Python解释器干了两件事情：

1. 在内存中创建了一个 'ABC' 的字符串；
2. 在内存中创建了一个名为 a 的变量，并把它指向 'ABC'。

也可以把一个变量 a 赋值给另一个变量 b，这个操作实际上是把变量 b 指向变量 a 所指向的数据，例如下面的代码：

```
# -*- coding: utf-8 -*-
a = 'ABC'
```

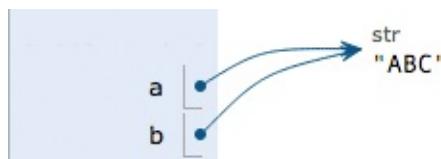
```
b = a
a = 'XYZ'
print(b)
```

最后一行打印出变量 `b` 的内容到底是 'ABC' 呢还是 'XYZ'？如果从数学意义上理解，就会错误地得出 `b` 和 `a` 相同，也应该是 'XYZ'，但实际上 `b` 的值是 'ABC'，让我们一行一行地执行代码，就可以看到到底发生了什么事：

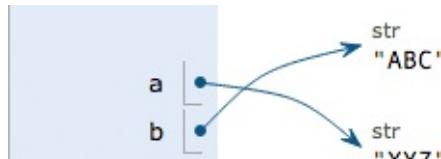
执行 `a = 'ABC'`，解释器创建了字符串 'ABC' 和变量 `a`，并把 `a` 指向 'ABC'：



执行 `b = a`，解释器创建了变量 `b`，并把 `b` 指向 `a` 指向的字符串 'ABC'：



执行 `a = 'XYZ'`，解释器创建了字符串 'XYZ'，并把 `a` 的指向改为 'XYZ'，但 `b` 并没有更改：



所以，最后打印变量 `b` 的结果自然是 'ABC' 了。

常量

所谓常量就是不能变的变量，比如常用的数学常数`π`就是一个常量。在Python中，通常用全部大写的变量名表示常量：

```
PI = 3.14159265359
```

但事实上 `PI` 仍然是一个变量，Python根本没有任何机制保证 `PI` 不会被改变，所以，用全部大写的变量名表示常量只是一个习惯上的用法，如果你一定要改变变量 `PI` 的值，也没人能拦住你。

最后解释一下整数的除法为什么也是精确的。在Python中，有两种除法，一种除法是 `/`：

```
>>> 10 / 3
3.3333333333333335
```

`/` 除法计算结果是浮点数，即使是两个整数恰好整除，结果也是浮点数：

```
>>> 9 / 3
3.0
```

还有一种除法是 `//`，称为地板除，两个整数的除法仍然是整数：

```
>>> 10 // 3
3
```

你没有看错，整数的地板除 `//` 永远是整数，即使除不尽。要做精确的除法，使用 `/` 就可以。

因为 `//` 除法只取结果的整数部分，所以 Python 还提供一个余数运算，可以得到两个整数相除的余数：

```
>>> 10 % 3
1
```

无论整数做 `//` 除法还是取余数，结果永远是整数，所以，整数运算结果永远是精确的。

练习

请打印出以下变量的值：

```
# -*- coding: utf-8 -*-
n = 123
f = 456.789
s1 = 'Hello, world'
s2 = 'Hello, \'Adam\''
s3 = r'Hello, "Bart"'
s4 = r'''Hello,
Lisa!'''
```

运行结果

```
123
456.789
Hello, world
Hello, 'Adam'
Hello, "Bart"
Hello,
Lisa!
```

小结

Python 支持多种数据类型，在计算机内部，可以把任何数据都看成一个“对象”，而变量就是在程序中用来指向这些数据对象的，对变量赋值就是把数据和变量给关联起来。

对变量赋值 `x = y` 是把变量 `x` 指向真正的对象，该对象是变量 `y` 所指向的。随后对变量 `y` 的赋值不影响变量 `x` 的指向。

注意：Python 的整数没有大小限制，而某些语言的整数根据其存储长度是有大小限制的，例如 Java 对 32 位整数的范围限制在 `-2147483648 ~ 2147483647`。

Python 的浮点数也没有大小限制，但是超出一定范围就直接表示为 `inf`（无限大）。

字符串和编码

字符编码

我们已经讲过了，字符串也是一种数据类型，但是，字符串比较特殊的是还有一个编码问题。

因为计算机只能处理数字，如果要处理文本，就必须先把文本转换为数字才能处理。最早的计算机在设计时采用8个比特（bit）作为一个字节（byte），所以，一个字节能表示的最大的整数就是255（二进制11111111=十进制255），如果要表示更大的整数，就必须用更多的字节。比如两个字节可以表示的最大整数是65535，4个字节可以表示的最大整数是4294967295。

由于计算机是美国人发明的，因此，最早只有127个字符被编码到计算机里，也就是大小写英文字母、数字和一些符号，这个编码表被称为ASCII编码，比如大写字母A的编码是65，小写字母z的编码是122。

但是要处理中文显然一个字节是不够的，至少需要两个字节，而且不能和ASCII编码冲突，所以，中国制定了GB2312编码，用来把中文编进去。

你可以想得到的是，全世界有上百种语言，日本把日文编到shift_JIS里，韩国把韩文编到Euc-kr里，各国有各国的标准，就会不可避免地出现冲突，结果就是，在多语言混合的文本中，显示出来会有乱码。



因此，Unicode应运而生。Unicode把所有语言都统一到一套编码里，这样就不会再有乱码问题了。

Unicode标准也在不断发展，但最常用的是用两个字节表示一个字符（如果要用到非常偏僻的字符，就需要4个字节）。现代操作系统和大多数编程语言都直接支持Unicode。

现在，捋一捋ASCII编码和Unicode编码的区别：ASCII编码是1个字节，而Unicode编码通常是2个字节。

字母A用ASCII编码是十进制的65，二进制的01000001；

字符@用ASCII编码是十进制的48，二进制的00110000，注意字符 '@' 和整数 0 是不同的；

汉字中已经超出了ASCII编码的范围，用Unicode编码是十进制的20013，二进制的01001110 00101101。

你可以猜测，如果把ASCII编码的A用Unicode编码，只需要在前面补0就可以，因此，A的Unicode编码是00000000 01000001。

新的问题又出现了：如果统一成Unicode编码，乱码问题从此消失了。但是，如果你写的文本基本上全部是英文的话，用Unicode编码比ASCII编码需要多一倍的存储空间，在存储和传输上就十分不划算。

所以，本着节约的精神，又出现了把Unicode编码转化为“可变长编码”的UTF-8编码。UTF-8编码把一个Unicode字符根据不同的数字大小编码成1-6个字节，常用的英文字母被编码成1个字节，汉字通常是3个字节，只有很生僻的字符才会被编码成4-6个字节。如果你要传输的文本包含大量英文字符，用UTF-8编码就能节省空间：

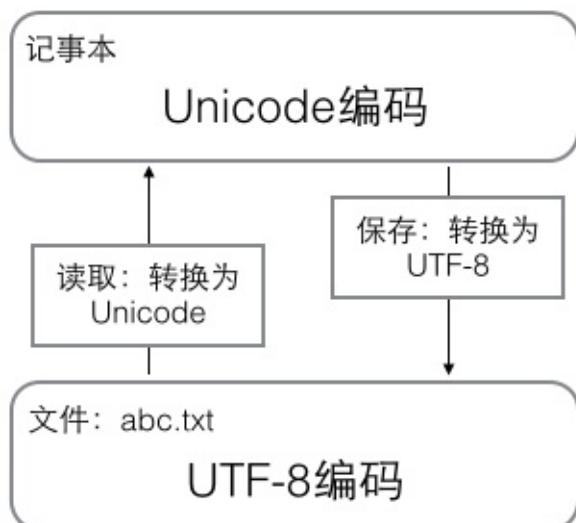
字符	ASCII	Unicode	UTF-8
A	01000001	00000000 01000001	01000001
中	x	01001110 00101101	11100100 10111000 10101101

从上面的表格还可以发现，UTF-8编码有一个额外的好处，就是ASCII编码实际上可以被看成是UTF-8编码的一部分，所以，大量只支持ASCII编码的历史遗留软件可以在UTF-8编码下继续工作。

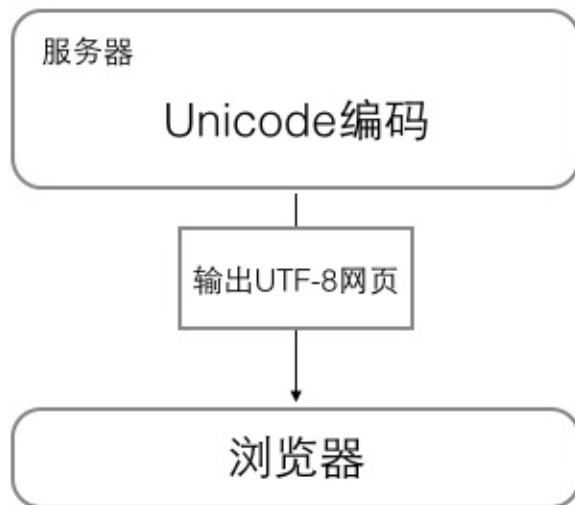
搞清楚了ASCII、Unicode和UTF-8的关系，我们就可以总结一下现在计算机系统通用的字符编码工作方式：

在计算机内存中，统一使用Unicode编码，当需要保存到硬盘或者需要传输的时候，就转换为UTF-8编码。

用记事本编辑的时候，从文件读取的UTF-8字符被转换为Unicode字符到内存里，编辑完成后，保存的时候再把Unicode转换为UTF-8保存到文件：



浏览网页的时候，服务器会把动态生成的Unicode内容转换为UTF-8再传输到浏览器：



所以你看到很多网页的源码上会有类似 `<meta charset="UTF-8" />` 的信息，表示该网页正是用的UTF-8编码。

Python的字符串

搞清楚了令人头疼的字符编码问题后，我们再来研究Python的字符串。

在最新的Python 3版本中，字符串是以Unicode编码的，也就是说，Python的字符串支持多语言，例如：

```
>>> print('包含中文的str')
包含中文的str
```

对于单个字符的编码，Python 提供了 `ord()` 函数获取字符的整数表示，`chr()` 函数把编码转换为对应的字符：

```
>>> ord('A')
65
>>> ord('中')
20013
>>> chr(66)
'B'
>>> chr(25991)
'文'
```

如果知道字符的整数编码，还可以用十六进制这么写 `str`：

```
>>> '\u4e2d\u6587'
'中文'
```

两种写法完全是等价的。

由于 Python 的字符串类型是 `str`，在内存中以 Unicode 表示，一个字符对应若干个字节。如果要在网络上传输，或者保存到磁盘上，就需要把 `str` 变为以字节为单位的 `bytes`。

Python 对 `bytes` 类型的数据用带 `b` 前缀的单引号或双引号表示：

```
x = b'ABC'
```

要注意区分 `'ABC'` 和 `b'ABC'`，前者是 `str`，后者虽然内容显示得和前者一样，但 `bytes` 的每个字符都只占用一个字节。

以 Unicode 表示的 `str` 通过 `encode()` 方法可以编码为指定的 `bytes`，例如：

```
>>> 'ABC'.encode('ascii')
b'ABC'
>>> '中文'.encode('utf-8')
b'\xe4\xb8\xad\xe6\x96\x87'
>>> '中文'.encode('ascii')
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
UnicodeEncodeError: 'ascii' codec can't encode characters in position 0-1: ordinal not in range(128)
```

纯英文的 `str` 可以用 `ASCII` 编码为 `bytes`，内容是一样的，含有中文的 `str` 可以用 `UTF-8` 编码为 `bytes`。含有中文的 `str` 无法用 `ASCII` 编码，因为中文编码的范围超过了 `ASCII` 编码的范围，Python 会报错。

在 `bytes` 中，无法显示为 ASCII 字符的字节，用 `\x##` 显示。

反过来，如果我们从网络或磁盘上读取了字节流，那么读到的数据就是 `bytes`。要把 `bytes` 变为 `str`，就需要用 `decode()` 方法：

```
>>> b'ABC'.decode('ascii')
'ABC'
>>> b'\xe4\xb8\xad\xe6\x96\x87'.decode('utf-8')
'中文'
```

如果 `bytes` 中包含无法解码的字节，`decode()` 方法会报错：

```
>>> b'\xe4\xb8\xad\xff'.decode('utf-8')
Traceback (most recent call last):
...
UnicodeDecodeError: 'utf-8' codec can't decode byte 0xff in position 3: invalid start byte
```

如果 `bytes` 中只有一小部分无效的字节，可以传入 `errors='ignore'` 忽略错误的字节：

```
>>> b'\xe4\xb8\xad\xff'.decode('utf-8', errors='ignore')
'中'
```

要计算 `str` 包含多少个字符，可以用 `len()` 函数：

```
>>> len('ABC')
3
>>> len('中文')
2
```

`len()` 函数计算的是 `str` 的字符数，如果换成 `bytes`，`len()` 函数就计算字节数：

```
>>> len(b'ABC')
3
>>> len(b'\xe4\xb8\xad\xe6\x96\x87')
6
>>> len('中文'.encode('utf-8'))
6
```

可见，1个中文字符经过UTF-8编码后通常会占用3个字节，而1个英文字符只占用1个字节。

在操作字符串时，我们经常遇到 `str` 和 `bytes` 的互相转换。为了避免乱码问题，应当始终坚持使用UTF-8编码对 `str` 和 `bytes` 进行转换。

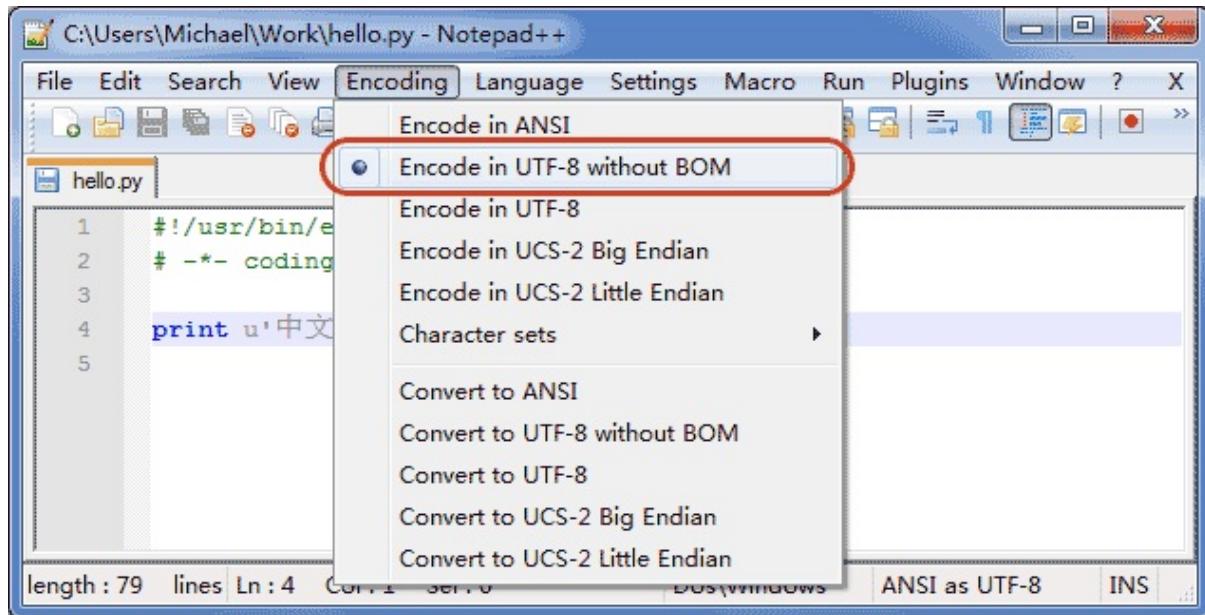
由于Python源代码也是一个文本文件，所以，当你的源代码中包含中文的时候，在保存源代码时，就需要务必指定保存为UTF-8编码。当Python解释器读取源代码时，为了让它按UTF-8编码读取，我们通常在文件开头写上这两行：

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
```

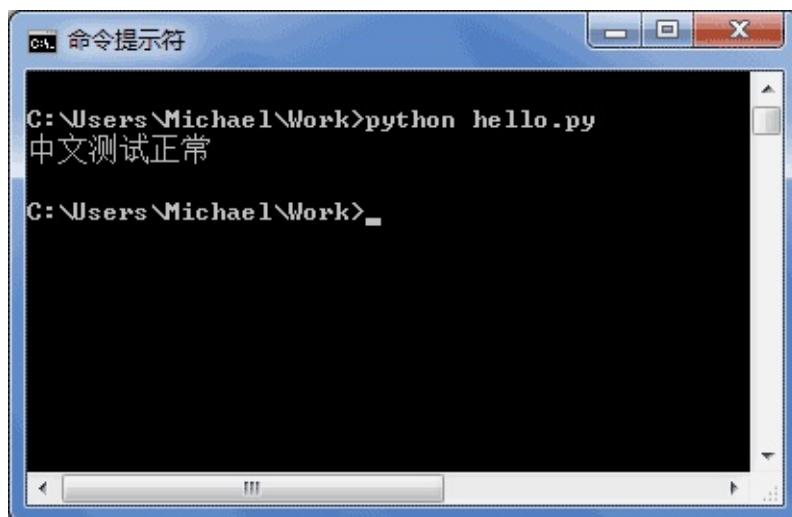
第一行注释是为了告诉Linux/OS X系统，这是一个Python可执行程序，Windows系统会忽略这个注释；

第二行注释是为了告诉Python解释器，按照UTF-8编码读取源代码，否则，你在源代码中写的中文输出可能会有乱码。

申明了UTF-8编码并不意味着你的 `.py` 文件就是UTF-8编码的，必须并且要确保文本编辑器正在使用UTF-8 without BOM编码：

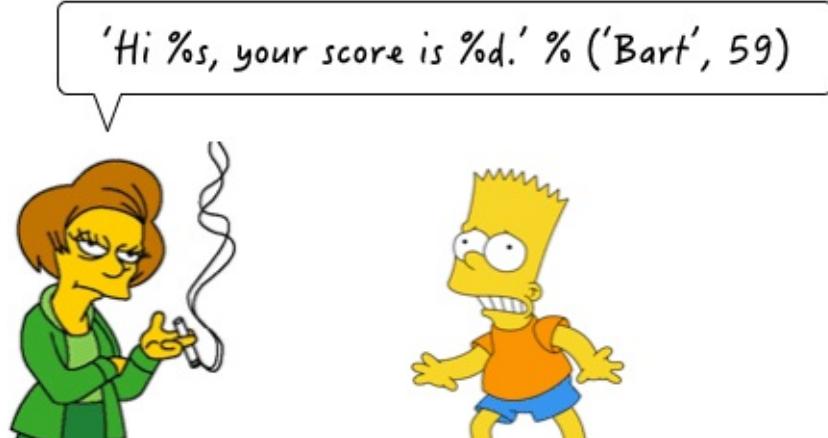


如果 .py 文件本身使用UTF-8编码，并且也申明了 `# -*- coding: utf-8 -*-`，打开命令提示符测试就可以正常显示中文：



格式化

最后一个常见的问题是输出格式化的字符串。我们经常会输出类似 '亲爱的xxx你好！你xx月的话费是xx，余额是xx' 之类的字符串，而xxx的内容都是根据变量变化的，所以，需要一种简便的格式化字符串的方式。



在Python中，采用的格式化方式和C语言是一致的，用%实现，举例如下：

```
>>> 'Hello, %s' % 'world'
'Hello, world'
>>> 'Hi, %s, you have $%d.' % ('Michael', 1000000)
'Hi, Michael, you have $1000000.'
```

你可能猜到了，%运算符就是用来格式化字符串的。在字符串内部，%s表示用字符串替换，%d表示用整数替换，有几个%占位符，后面就跟几个变量或者值，顺序要对应好。如果只有一个%?，括号可以省略。

常见的占位符有：

占位符	替换内容
%d	整数
%f	浮点数
%s	字符串
%x	十六进制整数

其中，格式化整数和浮点数还可以指定是否补0和整数与小数的位数：

```
# -*- coding: utf-8 -*-
print('%2d-%02d' % (3, 1))
print('.%2f' % 3.1415926)
```

运行结果

```
3-01
3.14
```

如果你不太确定应该用什么，%s永远起作用，它会把任何数据类型转换为字符串：

```
>>> 'Age: %s. Gender: %s' % (25, True)
'Age: 25. Gender: True'
```

有些时候，字符串里面的%是一个普通字符怎么办？这个时候就需要转义，用%%来表示一个%：

```
>>> 'growth rate: %d %%' % 7
'growth rate: 7 %'
```

format()

另一种格式化字符串的方法是使用字符串的 `format()` 方法，它会用传入的参数依次替换字符串内的占位符 `{0}`、`{1}`……，不过这种方式写起来比%要麻烦得多：

```
>>> 'Hello, {0}, 成绩提升了 {1:.1f}%' .format('小明', 17.125)
'Hello, 小明, 成绩提升了 17.1%'
```

练习

小明的成绩从去年的72分提升到了今年的85分，请计算小明成绩提升的百分点，并用字符串格式化显示出`'xx.x%'`，只保留小数点后1位：

```
# -*- coding: utf-8 -*-
s1 = 72
s2 = 85

r = (s2 - s1) / s1 * 100
print('%.1f%%' % r)
```

小结

Python 3 的字符串使用 Unicode，直接支持多语言。

当 `str` 和 `bytes` 互相转换时，需要指定编码。最常用的编码是 `UTF-8`。Python当然也支持其他编码方式，比如把 Unicode 编码成 `GB2312`：

```
>>> '中文'.encode('gb2312')
b'\xd6\xd0\xce\xc4'
```

但这种方式纯属自找麻烦，如果没有特殊业务要求，请牢记仅使用 `UTF-8` 编码。

格式化字符串的时候，可以用 Python 的交互式环境测试，方便快捷。

参考源码

[the_string.py](#)

使用list和tuple

list

Python内置的一种数据类型是列表：list。list是一种有序的集合，可以随时添加和删除其中的元素。

比如，列出班里所有同学的名字，就可以用一个list表示：

```
>>> classmates = ['Michael', 'Bob', 'Tracy']
>>> classmates
['Michael', 'Bob', 'Tracy']
```

变量 `classmates` 就是一个list。用 `len()` 函数可以获得list元素的个数：

```
>>> len(classmates)
3
```

用索引来访问list中每一个位置的元素，记得索引是从 0 开始的：

```
>>> classmates[0]
'Michael'
>>> classmates[1]
'Bob'
>>> classmates[2]
'Tracy'
>>> classmates[3]
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
IndexError: list index out of range
```

当索引超出了范围时，Python会报一个 `IndexError` 错误，所以，要确保索引不要越界，记得最后一个元素的索引是 `len(classmates) - 1`。

如果要取最后一个元素，除了计算索引位置外，还可以用 `-1` 做索引，直接获取最后一个元素：

```
>>> classmates[-1]
'Tracy'
```

以此类推，可以获取倒数第2个、倒数第3个：

```
>>> classmates[-2]
'Bob'
>>> classmates[-3]
'Michael'
>>> classmates[-4]
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
IndexError: list index out of range
```

当然，倒数第4个就越界了。

list是一个可变的有序表，所以，可以往list中追加元素到末尾：

```
>>> classmates.append('Adam')
>>> classmates
['Michael', 'Bob', 'Tracy', 'Adam']
```

也可以把元素插入到指定的位置，比如索引号为 1 的位置：

```
>>> classmates.insert(1, 'Jack')
>>> classmates
['Michael', 'Jack', 'Bob', 'Tracy', 'Adam']
```

要删除list末尾的元素，用 `pop()` 方法：

```
>>> classmates.pop()
'Adam'
>>> classmates
['Michael', 'Jack', 'Bob', 'Tracy']
```

要删除指定位置的元素，用 `pop(i)` 方法，其中 `i` 是索引位置：

```
>>> classmates.pop(1)
'Jack'
>>> classmates
['Michael', 'Bob', 'Tracy']
```

要把某个元素替换成别的元素，可以直接赋值给对应的索引位置：

```
>>> classmates[1] = 'Sarah'
>>> classmates
['Michael', 'Sarah', 'Tracy']
```

list里面的元素的数据类型也可以不同，比如：

```
>>> L = ['Apple', 123, True]
```

list元素也可以是另一个list，比如：

```
>>> s = ['python', 'java', ['asp', 'php'], 'scheme']
>>> len(s)
4
```

要注意 `s` 只有4个元素，其中 `s[2]` 又是一个list，如果拆开写就更容易理解了：

```
>>> p = ['asp', 'php']
>>> s = ['python', 'java', p, 'scheme']
```

要拿到 'php' 可以写 `p[1]` 或者 `s[2][1]`，因此 `s` 可以看成是一个二维数组，类似的还有三维、四维.....数组，不过很少用到。

如果一个list中一个元素也没有，就是一个空的list，它的长度为0：

```
>>> L = []
>>> len(L)
0
```

tuple

另一种有序列表叫元组：tuple。tuple和list非常类似，但是tuple一旦初始化就不能修改，比如同样是列出同学的名字：

```
>>> classmates = ('Michael', 'Bob', 'Tracy')
```

现在，`classmates`这个tuple不能变了，它也没有`append()`, `insert()`这样的方法。其他获取元素的方法和list是一样的，你可以正常地使用`classmates[0]`, `classmates[-1]`，但不能赋值成另外的元素。

不可变的tuple有什么意义？因为tuple不可变，所以代码更安全。如果可能，能用tuple代替list就尽量用tuple。

tuple的陷阱：当你定义一个tuple时，在定义的时候，tuple的元素就必须被确定下来，比如：

```
>>> t = (1, 2)
>>> t
(1, 2)
```

如果要定义一个空的tuple，可以写成`()`：

```
>>> t = ()
>>> t
()
```

但是，要定义一个只有1个元素的tuple，如果你这么定义：

```
>>> t = (1)
>>> t
1
```

定义的不是tuple，是`1`这个数！这是因为括号`()`既可以表示tuple，又可以表示数学公式中的小括号，这就产生了歧义，因此，Python规定，这种情况下，按小括号进行计算，计算结果自然是`1`。

所以，只有1个元素的tuple定义时必须加一个逗号`,`，来消除歧义：

```
>>> t = (1,)
>>> t
(1,)
```

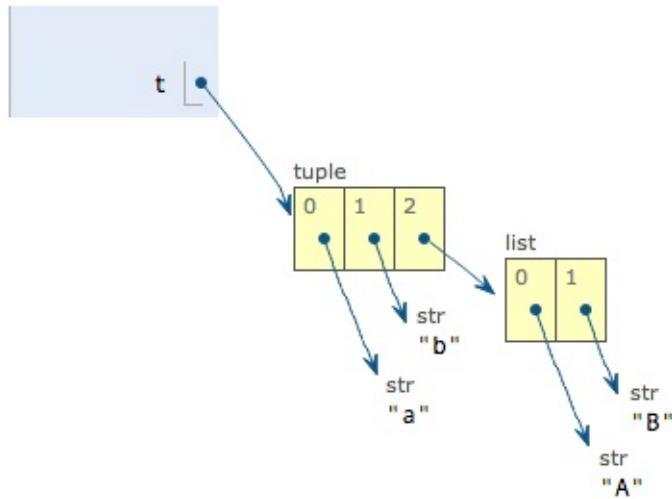
Python在显示只有1个元素的tuple时，也会加一个逗号`,`，以免你误解成数学计算意义上的括号。

最后来看一个“可变的”tuple：

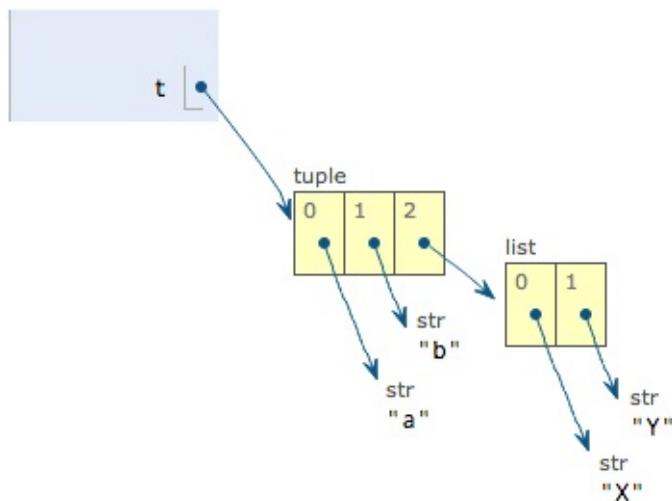
```
>>> t = ('a', 'b', ['A', 'B'])
>>> t[2][0] = 'X'
>>> t[2][1] = 'Y'
>>> t
('a', 'b', ['X', 'Y'])
```

这个tuple定义的时候有3个元素，分别是`'a'`, `'b'`和一个list。不是说tuple一旦定义后就不可变了吗？怎么后来又变了？

别急，我们先看看定义的时候tuple包含的3个元素：



当我们把list的元素 'A' 和 'B' 修改为 'X' 和 'Y' 后，tuple变为：



表面上看，tuple的元素确实变了，但其实变的不是tuple的元素，而是list的元素。tuple一开始指向的list并没有改成别的list，所以，tuple所谓的“不变”是说，tuple的每个元素，指向永远不变。即指向 'a'，就不能改成指向 'b'，指向一个list，就不能改成指向其他对象，但指向的这个list本身是可变的！

理解了“指向不变”后，要创建一个内容也不变的tuple怎么做？那就必须保证tuple的每一个元素本身也不能变。

练习

请用索引取出下面list的指定元素：

```

# -*- coding: utf-8 -*-
L = [
    ['Apple', 'Google', 'Microsoft'],
    ['Java', 'Python', 'Ruby', 'PHP'],
    ['Adam', 'Bart', 'Lisa']
]

# 打印Apple:
print(L[0][0])
# 打印Python:
print(L[1][1])
# 打印Lisa:
print(L[2][2])
  
```

运行结果

```
Apple  
Python  
Lisa
```

小结

list和tuple是Python内置的有序集合，一个可变，一个不可变。根据需要来选择使用它们。

参考源码

[the_list.py](#)

[the_tuple.py](#)

条件判断

条件判断

计算机之所以能做很多自动化的任务，因为它可以自己做条件判断。

比如，输入用户年龄，根据年龄打印不同的内容，在Python程序中，用 `if` 语句实现：

```
age = 20
if age >= 18:
    print('your age is', age)
    print('adult')
```

根据Python的缩进规则，如果 `if` 语句判断是 `True`，就把缩进的两行`print`语句执行了，否则，什么也不做。

也可以给 `if` 添加一个 `else` 语句，意思是，如果 `if` 判断是 `False`，不要执行 `if` 的内容，去把 `else` 执行了：

```
age = 3
if age >= 18:
    print('your age is', age)
    print('adult')
else:
    print('your age is', age)
    print('teenager')
```

注意不要少写了冒号 `:`。

当然上面的判断是很粗略的，完全可以用 `elif` 做更细致的判断：

```
age = 3
if age >= 18:
    print('adult')
elif age >= 6:
    print('teenager')
else:
    print('kid')
```

`elif` 是 `else if` 的缩写，完全可以有多个 `elif`，所以 `if` 语句的完整形式就是：

```
if <条件判断1>:
    <执行1>
elif <条件判断2>:
    <执行2>
elif <条件判断3>:
    <执行3>
else:
    <执行4>
```

`if` 语句执行有个特点，它是从上往下判断，如果在某个判断上是 `True`，把该判断对应的语句执行后，就忽略掉剩下的 `elif` 和 `else`，所以，请测试并解释为什么下面的程序打印的是 `teenager`：

```
age = 20
if age >= 6:
    print('teenager')
elif age >= 18:
    print('adult')
else:
    print('kid')
```

`if` 判断条件还可以简写，比如写：

```
if x:  
    print('True')
```

只要 `x` 是非零数值、非空字符串、非空list等，就判断为 `True`，否则为 `False`。

再议 `input`

最后看一个有问题的条件判断。很多同学会用 `input()` 读取用户的输入，这样可以自己输入，程序运行得更有意思：

```
birth = input('birth: ')  
if birth < 2000:  
    print('00前')  
else:  
    print('00后')
```

输入 `1982`，结果报错：

```
Traceback (most recent call last):  
  File "<stdin>", line 1, in <module>  
TypeError: unorderable types: str() > int()
```

这是因为 `input()` 返回的数据类型是 `str`，`str` 不能直接和整数比较，必须先把 `str` 转换成整数。Python提供了 `int()` 函数来完成这件事情：

```
s = input('birth: ')  
birth = int(s)  
if birth < 2000:  
    print('00前')  
else:  
    print('00后')
```

再次运行，就可以得到正确地结果。但是，如果输入 `abc` 呢？又会得到一个错误信息：

```
Traceback (most recent call last):  
  File "<stdin>", line 1, in <module>  
ValueError: invalid literal for int() with base 10: 'abc'
```

原来 `int()` 函数发现一个字符串并不是合法的数字时就会报错，程序就退出了。

如何检查并捕获程序运行期的错误呢？后面的错误和调试会讲到。

练习

小明身高1.75，体重80.5kg。请根据BMI公式（体重除以身高的平方）帮小明计算他的BMI指数，并根据BMI指数：

- 低于18.5：过轻
- 18.5-25：正常
- 25-28：过重
- 28-32：肥胖
- 高于32：严重肥胖

用 `if-elif` 判断并打印结果：

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

height = 1.75
weight = 80.5

bmi = weight / (height ** 2)
if bmi < 18.5:
    print('过轻')
elif bmi < 25:
    print('正常')
elif bmi < 28:
    print('过重')
elif bmi < 32:
    print('肥胖')
else:
    print('严重肥胖')
```

运行结果

过重

小结

条件判断可以让计算机自己做选择，Python的if...elif...else很灵活。

条件判断从上向下匹配，当满足条件时执行对应的块内语句，后续的elif和else都不再执行。

`if salary >= 10000:`



`elif salary >=5000:`



`else:`



参考源码

[do_if.py](#)

循环

循环

要计算 $1+2+3$ ，我们可以直接写表达式：

```
>>> 1 + 2 + 3
6
```

要计算 $1+2+3+\dots+10$ ，勉强也能写出来。

但是，要计算 $1+2+3+\dots+10000$ ，直接写表达式就不可能了。

为了让计算机能计算成千上万次的重复运算，我们就需要循环语句。

Python的循环有两种，一种是for...in循环，依次把list或tuple中的每个元素迭代出来，看例子：

```
names = ['Michael', 'Bob', 'Tracy']
for name in names:
    print(name)
```

执行这段代码，会依次打印 `names` 的每一个元素：

```
Michael
Bob
Tracy
```

所以 `for x in ...` 循环就是把每个元素代入变量 `x`，然后执行缩进块的语句。

再比如我们想计算 $1-10$ 的整数之和，可以用一个 `sum` 变量做累加：

```
sum = 0
for x in [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]:
    sum = sum + x
print(sum)
```

如果要计算 $1-100$ 的整数之和，从 1 写到 100 有点困难，幸好Python提供一个 `range()` 函数，可以生成一个整数序列，再通过 `list()` 函数可以转换为list。比如 `range(5)` 生成的序列是从 0 开始小于 5 的整数：

```
>>> list(range(5))
[0, 1, 2, 3, 4]
```

`range(101)` 就可以生成 $0-100$ 的整数序列，计算如下：

```
# -*- coding: utf-8 -*-
sum = 0
for x in range(101):
    sum = sum + x
print(sum)
```

运行结果

```
5050
```

请自行运行上述代码，看看结果是不是当年高斯同学心算出的5050。

第二种循环是while循环，只要条件满足，就不断循环，条件不满足时退出循环。比如我们要计算100以内所有奇数之和，可以用while循环实现：

```
sum = 0
n = 99
while n > 0:
    sum = sum + n
    n = n - 2
print(sum)
```

在循环内部变量 n 不断自减，直到变为 -1 时，不再满足while条件，循环退出。

练习

请利用循环依次对list中的每个名字打印出 Hello, xxx! :

```
# -*- coding: utf-8 -*-
L = ['Bart', 'Lisa', 'Adam']

for name in L:
    print('Hello, %s!' % name)
```

运行结果

```
Hello, Bart!
Hello, Lisa!
Hello, Adam!
```

break

在循环中， break 语句可以提前退出循环。例如，本来要循环打印1~100的数字：

```
n = 1
while n <= 100:
    print(n)
    n = n + 1
print('END')
```

上面的代码可以打印出1~100。

如果要提前结束循环，可以用 break 语句：

```
n = 1
while n <= 100:
    if n > 10: # 当n = 11时，条件满足，执行break语句
        break # break语句会结束当前循环
    print(n)
    n = n + 1
print('END')
```

执行上面的代码可以看到，打印出1~10后，紧接着打印 END ，程序结束。

可见 break 的作用是提前结束循环。

continue

在循环过程中，也可以通过 `continue` 语句，跳过当前的这次循环，直接开始下一次循环。

```
n = 0
while n < 10:
    n = n + 1
    print(n)
```

上面的程序可以打印出1~10。但是，如果我们想只打印奇数，可以用 `continue` 语句跳过某些循环：

```
n = 0
while n < 10:
    n = n + 1
    if n % 2 == 0: # 如果n是偶数，执行continue语句
        continue # continue语句会直接继续下一轮循环，后续的print()语句不会执行
    print(n)
```

执行上面的代码可以看到，打印的不再是1~10，而是1, 3, 5, 7, 9。

可见 `continue` 的作用是提前结束本轮循环，并直接开始下一轮循环。

小结

循环是让计算机做重复任务的有效方法。

`break` 语句可以在循环过程中直接退出循环，而 `continue` 语句可以提前结束本轮循环，并直接开始下一轮循环。这两个语句通常都必须配合 `if` 语句使用。

要特别注意，不要滥用 `break` 和 `continue` 语句。`break` 和 `continue` 会造成代码执行逻辑分叉过多，容易出错。大多数循环并不需要用到 `break` 和 `continue` 语句，上面的两个例子，都可以通过改写循环条件或者修改循环逻辑，去掉 `break` 和 `continue` 语句。

有些时候，如果代码写得有问题，会让程序陷入“死循环”，也就是永远循环下去。这时可以用 `ctrl+C` 退出程序，或者强制结束Python进程。

请试写一个死循环程序。

参考源码

[do_for.py](#)

[do_while.py](#)

使用dict和set

dict

Python内置了字典：dict的支持，dict全称dictionary，在其他语言中也称为map，使用键-值（key-value）存储，具有极快的查找速度。

举个例子，假设要根据同学的名字查找对应的成绩，如果用list实现，需要两个list：

```
names = ['Michael', 'Bob', 'Tracy']
scores = [95, 75, 85]
```

给定一个名字，要查找对应的成绩，就先要在names中找到对应的位置，再从scores取出对应的成绩，list越长，耗时越长。

如果用dict实现，只需要一个“名字”-“成绩”的对照表，直接根据名字查找成绩，无论这个表有多大，查找速度都不会变慢。用Python写一个dict如下：

```
>>> d = {'Michael': 95, 'Bob': 75, 'Tracy': 85}
>>> d['Michael']
95
```

为什么dict查找速度这么快？因为dict的实现原理和查字典是一样的。假设字典包含了1万个汉字，我们要查某一个字，一个办法是把字典从第一页往后翻，直到找到我们想要的字为止，这种方法就是在list中查找元素的方法，list越大，查找越慢。

第二种方法是先在字典的索引表里（比如部首表）查这个字对应的页码，然后直接翻到该页，找到这个字。无论找哪个字，这种查找速度都非常快，不会随着字典大小的增加而变慢。

dict就是第二种实现方式，给定一个名字，比如 'Michael'，dict在内部就可以直接计算出 Michael 对应的存放成绩的“页码”，也就是 95 这个数字存放的内存地址，直接取出来，所以速度非常快。

你可以猜到，这种key-value存储方式，在放进去的时候，必须根据key算出value的存放位置，这样，取的时候才能根据key直接拿到value。

把数据放入dict的方法，除了初始化时指定外，还可以通过key放入：

```
>>> d['Adam'] = 67
>>> d['Adam']
67
```

由于一个key只能对应一个value，所以，多次对一个key放入value，后面的值会把前面的值冲掉：

```
>>> d['Jack'] = 90
>>> d['Jack']
90
>>> d['Jack'] = 88
>>> d['Jack']
88
```

如果key不存在，dict就会报错：

```
>>> d['Thomas']
Traceback (most recent call last):
File "<stdin>", line 1, in <module>
```

```
KeyError: 'Thomas'
```

要避免key不存在的错误，有两种办法，一是通过 `in` 判断key是否存在：

```
>>> 'Thomas' in d
False
```

二是通过dict提供的 `get()` 方法，如果key不存在，可以返回 `None`，或者自己指定的value：

```
>>> d.get('Thomas')
>>> d.get('Thomas', -1)
-1
```

注意：返回 `None` 的时候Python的交互环境不显示结果。

要删除一个key，用 `pop(key)` 方法，对应的value也会从dict中删除：

```
>>> d.pop('Bob')
75
>>> d
{'Michael': 95, 'Tracy': 85}
```

请务必注意，dict内部存放的顺序和key放入的顺序是没有关系的。

和list比较，dict有以下几个特点：

1. 查找和插入的速度极快，不会随着key的增加而变慢；
2. 需要占用大量的内存，内存浪费多。

而list相反：

1. 查找和插入的时间随着元素的增加而增加；
2. 占用空间小，浪费内存很少。

所以，dict是用空间来换取时间的一种方法。

dict可以用在需要高速查找的很多地方，在Python代码中几乎无处不在，正确使用dict非常重要，需要牢记的第一条就是dict的key必须是不可变对象。

这是因为dict根据key来计算value的存储位置，如果每次计算相同的key得出的结果不同，那dict内部就完全混乱了。这个通过key计算位置的算法称为哈希算法（Hash）。

要保证hash的正确性，作为key的对象就不能变。在Python中，字符串、整数等都是不可变的，因此，可以放心地作为key。而list是可变的，就不能作为key：

```
>>> key = [1, 2, 3]
>>> d[key] = 'a list'
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: unhashable type: 'list'
```

set

set和dict类似，也是一组key的集合，但不存储value。由于key不能重复，所以，在set中，没有重复的key。

要创建一个set，需要提供一个list作为输入集合：

```
>>> s = set([1, 2, 3])
```

```
>>> s
{1, 2, 3}
```

注意，传入的参数 [1, 2, 3] 是一个list，而显示的 {1, 2, 3} 只是告诉你这个set内部有1, 2, 3这3个元素，显示的顺序也不表示set是有序的。

重复元素在set中自动被过滤：

```
>>> s = set([1, 1, 2, 2, 3, 3])
>>> s
{1, 2, 3}
```

通过 add(key) 方法可以添加元素到set中，可以重复添加，但不会有效果：

```
>>> s.add(4)
>>> s
{1, 2, 3, 4}
>>> s.add(4)
>>> s
{1, 2, 3, 4}
```

通过 remove(key) 方法可以删除元素：

```
>>> s.remove(4)
>>> s
{1, 2, 3}
```

set可以看成数学意义上的无序和无重复元素的集合，因此，两个set可以做数学意义上的交集、并集等操作：

```
>>> s1 = set([1, 2, 3])
>>> s2 = set([2, 3, 4])
>>> s1 & s2
{2, 3}
>>> s1 | s2
{1, 2, 3, 4}
```

set和dict的唯一区别仅在于没有存储对应的value，但是，set的原理和dict一样，所以，同样不可以放入可变对象，因为无法判断两个可变对象是否相等，也就无法保证set内部“不会有重复元素”。试试把list放入set，看看是否会报错。

再议不可变对象

上面我们讲了，str是不变对象，而list是可变对象。

对于可变对象，比如list，对list进行操作，list内部的内容是会变化的，比如：

```
>>> a = ['c', 'b', 'a']
>>> a.sort()
>>> a
['a', 'b', 'c']
```

而对于不可变对象，比如str，对str进行操作呢：

```
>>> a = 'abc'
>>> a.replace('a', 'A')
'Abc'
>>> a
'abc'
```

虽然字符串有个 `replace()` 方法，也确实变出了 `'Abc'`，但变量 `a` 最后仍是 `'abc'`，应该怎么理解呢？

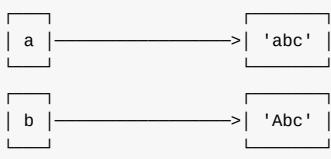
我们先把代码改成下面这样：

```
>>> a = 'abc'
>>> b = a.replace('a', 'A')
>>> b
'Abc'
>>> a
'abc'
```

要始终牢记的是，`a` 是变量，而 `'abc'` 才是字符串对象！有些时候，我们经常说，对象 `a` 的内容是 `'abc'`，但其实是指，`a` 本身是一个变量，它指向的对象的内容才是 `'abc'`：



当我们调用 `a.replace('a', 'A')` 时，实际上调用方法 `replace` 是作用在字符串对象 `'abc'` 上的，而这个方法虽然名字叫 `replace`，但却没有改变字符串 `'abc'` 的内容。相反，`replace` 方法创建了一个新字符串 `'Abc'` 并返回，如果我们用变量 `b` 指向该新字符串，就容易理解了，变量 `a` 仍指向原有的字符串 `'abc'`，但变量 `b` 却指向新字符串 `'Abc'` 了：



所以，对于不变对象来说，调用对象自身的任意方法，也不会改变该对象自身的内容。相反，这些方法会创建新的对象并返回，这样，就保证了不可变对象本身永远是不可变的。

小结

使用key-value存储结构的dict在Python中非常有用，选择不可变对象作为key很重要，最常用的key是字符串。

tuple虽然是不变对象，但试试把 `(1, 2, 3)` 和 `(1, [2, 3])` 放入dict或set中，并解释结果。

参考源码

[the_dict.py](#)

[the_set.py](#)

函数

我们知道圆的面积计算公式为：

$$S = \pi r^2$$

当我们知道半径 r 的值时，就可以根据公式计算出面积。假设我们需要计算3个不同大小的圆的面积：

```
r1 = 12.34
r2 = 9.08
r3 = 73.1
s1 = 3.14 * r1 * r1
s2 = 3.14 * r2 * r2
s3 = 3.14 * r3 * r3
```

当代码出现有规律的重复的时候，你就需要当心了，每次写 $3.14 * x * x$ 不仅很麻烦，而且，如果要把 3.14 改成 3.14159265359 的时候，得全部替换。

有了函数，我们就不再每次写 $s = 3.14 * x * x$ ，而是写成更有意义的函数调用 $s = area_of_circle(x)$ ，而函数 $area_of_circle$ 本身只需要写一次，就可以多次调用。

基本上所有的高级语言都支持函数，Python也不例外。Python不但能非常灵活地定义函数，而且本身内置了很多有用的函数，可以直接调用。

抽象

抽象是数学中非常常见的概念。举个例子：

计算数列的和，比如： $1 + 2 + 3 + \dots + 100$ ，写起来十分不方便，于是数学家发明了求和符号 Σ ，可以把 $1 + 2 + 3 + \dots + 100$ 记作：

$$\sum_{n=1}^{100} n$$

这种抽象记法非常强大，因为我们看到 Σ 就可以理解成求和，而不是还原成低级的加法运算。

而且，这种抽象记法是可扩展的，比如：

$$\sum_{n=1}^{100} (n^2 + 1)$$

还原成加法运算就变成了：

$$(1 \times 1 + 1) + (2 \times 2 + 1) + (3 \times 3 + 1) + \dots + (100 \times 100 + 1)$$

可见，借助抽象，我们才能不关心底层的具体计算过程，而直接在更高的层次上思考问题。

写计算机程序也是一样，函数就是最基本的一种代码抽象的方式。

调用函数

Python内置了很多有用的函数，我们可以直接调用。

要调用一个函数，需要知道函数的名称和参数，比如求绝对值的函数 `abs`，只有一个参数。可以直接从Python的官方网站查看文档：

<http://docs.python.org/3/library/functions.html#abs>

也可以在交互式命令行通过 `help(abs)` 查看 `abs` 函数的帮助信息。

调用 `abs` 函数：

```
>>> abs(100)
100
>>> abs(-20)
20
>>> abs(12.34)
12.34
```

调用函数的时候，如果传入的参数数量不对，会报 `TypeError` 的错误，并且Python会明确地告诉你：`abs()` 有且仅有1个参数，但给出了两个：

```
>>> abs(1, 2)
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: abs() takes exactly one argument (2 given)
```

如果传入的参数数量是对的，但参数类型不能被函数所接受，也会报 `TypeError` 的错误，并且给出错误信息：`str` 是错误的参数类型：

```
>>> abs('a')
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: bad operand type for abs(): 'str'
```

而 `max` 函数 `max()` 可以接收任意多个参数，并返回最大的那个：

```
>>> max(1, 2)
2
>>> max(2, 3, 1, -5)
3
```

数据类型转换

Python内置的常用函数还包括数据类型转换函数，比如 `int()` 函数可以把其他数据类型转换为整数：

```
>>> int('123')
123
>>> int(12.34)
12
>>> float('12.34')
12.34
>>> str(1.23)
'1.23'
>>> str(100)
```

```
'100'
>>> bool(1)
True
>>> bool('')
False
```

函数名其实就是指向一个函数对象的引用，完全可以把函数名赋给一个变量，相当于给这个函数起了一个“别名”：

```
>>> a = abs # 变量a指向abs函数
>>> a(-1) # 所以也可以通过a调用abs函数
1
```

练习

请利用Python内置的 `hex()` 函数把一个整数转换成十六进制表示的字符串：

```
# -*- coding: utf-8 -*-
n1 = 255
n2 = 1000

print(hex(n1))
print(hex(n2))
```

运行结果

```
0xff
0x3e8
```

小结

调用Python的函数，需要根据函数定义，传入正确的参数。如果函数调用出错，一定要学会看错误信息，所以英文很重要！

参考源码

[call_func.py](#)

定义函数

在Python中，定义一个函数要使用 `def` 语句，依次写出函数名、括号、括号中的参数和冒号`:`，然后，在缩进块中编写函数体，函数的返回值用 `return` 语句返回。

我们以自定义一个求绝对值的 `my_abs` 函数为例：

```
# -*- coding: utf-8 -*-
def my_abs(x):
    if x >= 0:
        return x
    else:
        return -x

print(my_abs(-99))
```

运行结果

```
99
```

请自行测试并调用 `my_abs` 看看返回结果是否正确。

请注意，函数体内部的语句在执行时，一旦执行到 `return` 时，函数就执行完毕，并将结果返回。因此，函数内部通过条件判断和循环可以实现非常复杂的逻辑。

如果没有 `return` 语句，函数执行完毕后也会返回结果，只是结果为 `None`。`return None` 可以简写为 `return`。

在Python交互环境中定义函数时，注意Python会出现`...`的提示。函数定义结束后需要按两次回车重新回到`>>>`提示符下：

```
|Command Prompt - python
|>>> def my_abs(x):
|...     if x >= 0:
|...         return x
|...     else:
|...         return -x
|...
|>>> my_abs(-9)
|9
|>>> _
```

如果你已经把 `my_abs()` 的函数定义保存为 `abstest.py` 文件了，那么，可以在该文件的当前目录下启动Python解释器，用 `from abstest import my_abs` 来导入 `my_abs()` 函数，注意 `abstest` 是文件名（不含 `.py` 扩展名）：

```
|Command Prompt - python
|>>> from abstest import my_abs
|>>> my_abs(-9)
|9
|>>> _
```



`import` 的用法在后续[模块](#)一节中会详细介绍。

空函数

如果想定义一个什么事也不做的空函数，可以用 `pass` 语句：

```
def nop():
    pass
```

`pass` 语句什么都不做，那有什么用？实际上 `pass` 可以用来作为占位符，比如现在还没想好怎么写函数的代码，就可以先放一个 `pass`，让代码能运行起来。

`pass` 还可以用在其他语句里，比如：

```
if age >= 18:
    pass
```

缺少了 `pass`，代码运行就会有语法错误。

参数检查

调用函数时，如果参数个数不对，Python 解释器会自动检查出来，并抛出 `TypeError`：

```
>>> my_abs(1, 2)
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: my_abs() takes 1 positional argument but 2 were given
```

但是如果参数类型不对，Python 解释器就无法帮我们检查。试试 `my_abs` 和内置函数 `abs` 的差别：

```
>>> my_abs('A')
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
  File "<stdin>", line 2, in my_abs
TypeError: unorderable types: str() >= int()
>>> abs('A')
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: bad operand type for abs(): 'str'
```

当传入了不恰当的参数时，内置函数 `abs` 会检查出参数错误，而我们定义的 `my_abs` 没有参数检查，会导致 `if` 语句出错，出错信息和 `abs` 不一样。所以，这个函数定义不够完善。

让我们修改一下 `my_abs` 的定义，对参数类型做检查，只允许整数和浮点数类型的参数。数据类型检查可以用内置函数 `isinstance()` 实现：

```
def my_abs(x):
    if not isinstance(x, (int, float)):
        raise TypeError('bad operand type')
    if x >= 0:
```

```

        return x
    else:
        return -x

```

添加了参数检查后，如果传入错误的参数类型，函数就可以抛出一个错误：

```

>>> my_abs('A')
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
  File "<stdin>", line 3, in my_abs
TypeError: bad operand type

```

错误和异常处理将在后续讲到。

返回多个值

函数可以返回多个值吗？答案是肯定的。

比如在游戏中经常需要从一个点移动到另一个点，给出坐标、位移和角度，就可以计算出新的新的坐标：

```

import math

def move(x, y, step, angle=0):
    nx = x + step * math.cos(angle)
    ny = y - step * math.sin(angle)
    return nx, ny

```

`import math` 语句表示导入 `math` 包，并允许后续代码引用 `math` 包里的 `sin`、`cos` 等函数。

然后，我们就可以同时获得返回值：

```

>>> x, y = move(100, 100, 60, math.pi / 6)
>>> print(x, y)
151.96152422706632 70.0

```

但其实这只不过是一种假象，Python 函数返回的仍然是单一值：

```

>>> r = move(100, 100, 60, math.pi / 6)
>>> print(r)
(151.96152422706632, 70.0)

```

原来返回值是一个tuple！但是，在语法上，返回一个tuple可以省略括号，而多个变量可以同时接收一个tuple，按位置赋给对应的值，所以，Python 的函数返回多值其实就是返回一个tuple，但写起来更方便。

小结

定义函数时，需要确定函数名和参数个数；

如果有必要，可以先对参数的数据类型做检查；

函数体内部可以用 `return` 随时返回函数结果；

函数执行完毕也没有 `return` 语句时，自动 `return None`。

函数可以同时返回多个值，但其实就是一个tuple。

练习

请定义一个函数 `quadratic(a, b, c)`，接收3个参数，返回一元二次方程：

$$ax^2 + bx + c = 0$$

的两个解。

提示：计算平方根可以调用 `math.sqrt()` 函数：

```
>>> import math
>>> math.sqrt(2)
1.4142135623730951
```

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

def quadratic(a, b, c):
    if a == 0:
        return ('非一元二次方程式')
    else:
        r = b*b - 4*a*c
        if r < 0:
            return ('无实根')
        else:
            x1 = (-b + math.sqrt(r))/(2*a)
            x2 = (-b - math.sqrt(r))/(2*a)
            return x1, x2

# 测试:
print('quadratic(2, 3, 1) =', quadratic(2, 3, 1))
print('quadratic(1, 3, -4) =', quadratic(1, 3, -4))

if quadratic(2, 3, 1) != (-0.5, -1.0):
    print('测试失败!')
elif quadratic(1, 3, -4) != (1.0, -4.0):
    print('测试失败!')
else:
    print('测试成功!')
```

参考源码

[def_func.py](#)

函数的参数

定义函数的时候，我们把参数的名字和位置确定下来，函数的接口定义就完成了。对于函数的调用者来说，只需要知道如何传递正确的参数，以及函数将返回什么样的值就够了，函数内部的复杂逻辑被封装起来，调用者无需了解。

Python 的函数定义非常简单，但灵活度却非常大。除了正常定义的必选参数外，还可以使用默认参数、可变参数和关键字参数，使得函数定义出来的接口，不但能处理复杂的参数，还可以简化调用者的代码。

位置参数

我们先写一个计算 x^2 的函数：

```
def power(x):
    return x * x
```

对于 `power(x)` 函数，参数 `x` 就是一个位置参数。

当我们调用 `power` 函数时，必须传入有且仅有的一个参数 `x`：

```
>>> power(5)
25
>>> power(15)
225
```

现在，如果我们要计算 x^3 怎么办？可以再定义一个 `power3` 函数，但是如果要计算 x^4 、 x^5 ……怎么办？我们不可能定义无限多个函数。

你也许想到了，可以把 `power(x)` 修改为 `power(x, n)`，用来计算 x^n ，说干就干：

```
def power(x, n):
    s = 1
    while n > 0:
        n = n - 1
        s = s * x
    return s
```

对于这个修改后的 `power(x, n)` 函数，可以计算任意 n 次方：

```
>>> power(5, 2)
25
>>> power(5, 3)
125
```

修改后的 `power(x, n)` 函数有两个参数：`x` 和 `n`，这两个参数都是位置参数，调用函数时，传入的两个值按照位置顺序依次赋给参数 `x` 和 `n`。

默认参数

新的 `power(x, n)` 函数定义没有问题，但是，旧的调用代码失败了，原因是我们在增加了一个参数，导致旧的代码因为缺少一个参数而无法正常调用：

```
>>> power(5)
Traceback (most recent call last):
File "<stdin>", line 1, in <module>
```

```
TypeError: power() missing 1 required positional argument: 'n'
```

Python 的错误信息很明确：调用函数 `power()` 缺少了一个位置参数 `n`。

这个时候，默认参数就派上用场了。由于我们经常计算 x^2 ，所以，完全可以把第二个参数 `n` 的默认值设定为 2：

```
def power(x, n=2):
    s = 1
    while n > 0:
        n = n - 1
        s = s * x
    return s
```

这样，当我们调用 `power(5)` 时，相当于调用 `power(5, 2)`：

```
>>> power(5)
25
>>> power(5, 2)
25
```

而对于 `n > 2` 的其他情况，就必须明确地传入 `n`，比如 `power(5, 3)`。

从上面的例子可以看出，默认参数可以简化函数的调用。设置默认参数时，有几点要注意：

- 一是必选参数在前，默认参数在后，否则 Python 的解释器会报错（思考一下为什么默认参数不能放在必选参数前面）；
- 二是如何设置默认参数。

当函数有多个参数时，把变化大的参数放前面，变化小的参数放后面。变化小的参数就可以作为默认参数。

使用默认参数有什么好处？最大的好处是能降低调用函数的难度。

举个例子，我们写个一年级小学生注册的函数，需要传入 `name` 和 `gender` 两个参数：

```
def enroll(name, gender):
    print('name:', name)
    print('gender:', gender)
```

这样，调用 `enroll()` 函数只需要传入两个参数：

```
>>> enroll('Sarah', 'F')
name: Sarah
gender: F
```

如果要继续传入年龄、城市等信息怎么办？这样会使得调用函数的复杂度大大增加。

我们可以把年龄和城市设为默认参数：

```
def enroll(name, gender, age=6, city='Beijing'):
    print('name:', name)
    print('gender:', gender)
    print('age:', age)
    print('city:', city)
```

这样，大多数学生注册时不需要提供年龄和城市，只提供必须的两个参数：

```
>>> enroll('Sarah', 'F')
name: Sarah
gender: F
```

```
age: 6
city: Beijing
```

只有与默认参数不符的学生才需要提供额外的信息：

```
enroll('Bob', 'M', 7)
enroll('Adam', 'M', city='Tianjin')
```

可见，默认参数降低了函数调用的难度，而一旦需要更复杂的调用时，又可以传递更多的参数来实现。无论是简单调用还是复杂调用，函数只需要定义一个。

有多个默认参数时，调用的时候，既可以按顺序提供默认参数，比如调用 `enroll('Bob', 'M', 7)`，意思是，除了 `name`，`gender` 这两个参数外，最后1个参数应用在参数 `age` 上，`city` 参数由于没有提供，仍然使用默认值。

也可以不按顺序提供部分默认参数。当不按顺序提供部分默认参数时，需要把参数名写上。比如调用 `enroll('Adam', 'M', city='Tianjin')`，意思是，`city` 参数用传进去的值，其他默认参数继续使用默认值。

默认参数很有用，但使用不当，也会掉坑里。默认参数有个最大的坑，演示如下：

先定义一个函数，传入一个list，添加一个END再返回：

```
def add_end(L=[]):
    L.append('END')
    return L
```

当你正常调用时，结果似乎不错：

```
>>> add_end([1, 2, 3])
[1, 2, 3, 'END']
>>> add_end(['x', 'y', 'z'])
['x', 'y', 'z', 'END']
```

当你使用默认参数调用时，一开始结果也是对的：

```
>>> add_end()
['END']
```

但是，再次调用 `add_end()` 时，结果就不对了：

```
>>> add_end()
['END', 'END']
>>> add_end()
['END', 'END', 'END']
```

很多初学者很疑惑，默认参数是 `[]`，但是函数似乎每次都“记住了”上次添加了 `'END'` 后的list。

原因解释如下：

Python函数在定义的时候，默认参数 `L` 的值就被计算出来了，即 `[]`，因为默认参数 `L` 也是一个变量，它指向对象 `[]`，每次调用该函数，如果改变了 `L` 的内容，则下次调用时，默认参数的内容就变了，不再是函数定义时的 `[]` 了。

定义默认参数要牢记一点：默认参数必须指向不变对象！

要修改上面的例子，我们可以用 `None` 这个不变对象来实现：

```
def add_end(L=None):
    if L is None:
```

```
L = []
L.append('END')
return L
```

现在，无论调用多少次，都不会有问题：

```
>>> add_end()
['END']
>>> add_end()
['END']
```

为什么要设计 `str`、`None` 这样的不变对象呢？因为不变对象一旦创建，对象内部的数据就不能修改，这样就减少了由于修改数据导致的错误。此外，由于对象不变，多任务环境下同时读取对象不需要加锁，同时读一点问题都没有。我们在编写程序时，如果可以设计一个不变对象，那就尽量设计成不变对象。

可变参数

在Python函数中，还可以定义可变参数。顾名思义，可变参数就是传入的参数个数是可变的，可以是1个、2个到任意个，还可以是0个。

我们以数学题为例子，给定一组数字a, b, c.....，请计算 $a^2 + b^2 + c^2 + \dots$ 。

要定义出这个函数，我们必须确定输入的参数。由于参数个数不确定，我们首先想到可以把a, b, c.....作为一个list或tuple传进来，这样，函数可以定义如下：

```
def calc(numbers):
    sum = 0
    for n in numbers:
        sum = sum + n * n
    return sum
```

但是调用的时候，需要先组装出一个list或tuple：

```
>>> calc([1, 2, 3])
14
>>> calc((1, 3, 5, 7))
84
```

如果利用可变参数，调用函数的方式可以简化成这样：

```
>>> calc(1, 2, 3)
14
>>> calc(1, 3, 5, 7)
84
```

所以，我们把函数的参数改为可变参数：

```
def calc(*numbers):
    sum = 0
    for n in numbers:
        sum = sum + n * n
    return sum
```

定义可变参数和定义一个list或tuple参数相比，仅仅在参数前面加了一个`*`号。在函数内部，参数 `numbers` 接收到的是一个tuple，因此，函数代码完全不变。但是，调用该函数时，可以传入任意个参数，包括0个参数：

```
>>> calc(1, 2)
```

```
5
>>> calc()
0
```

如果已经有一个list或者tuple，要调用一个可变参数怎么办？可以这样做：

```
>>> nums = [1, 2, 3]
>>> calc(nums[0], nums[1], nums[2])
14
```

这种写法当然是可行的，问题是太繁琐，所以Python允许你在list或tuple前面加一个`*`号，把list或tuple的元素变成可变参数传进去：

```
>>> nums = [1, 2, 3]
>>> calc(*nums)
14
```

`*nums` 表示把 `nums` 这个list的所有元素作为可变参数传进去。这种写法相当有用，而且很常见。

关键字参数

可变参数允许你传入0个或任意个参数，这些可变参数在函数调用时自动组装为一个tuple。而关键字参数允许你传入0个或任意个含参数名的参数，这些关键字参数在函数内部自动组装为一个dict。请看示例：

```
def person(name, age, **kw):
    print('name:', name, 'age:', age, 'other:', kw)
```

函数 `person` 除了必选参数 `name` 和 `age` 外，还接受关键字参数 `kw`。在调用该函数时，可以只传入必选参数：

```
>>> person('Michael', 30)
name: Michael age: 30 other: {}
```

也可以传入任意个数的关键字参数：

```
>>> person('Bob', 35, city='Beijing')
name: Bob age: 35 other: {'city': 'Beijing'}
>>> person('Adam', 45, gender='M', job='Engineer')
name: Adam age: 45 other: {'gender': 'M', 'job': 'Engineer'}
```

关键字参数有什么用？它可以扩展函数的功能。比如，在 `person` 函数里，我们保证能接收到 `name` 和 `age` 这两个参数，但是，如果调用者愿意提供更多的参数，我们也能收到。试想你正在做一个用户注册的功能，除了用户名和年龄是必填项外，其他都是可选项，利用关键字参数来定义这个函数就能满足注册的需求。

和可变参数类似，也可以先组装出一个dict，然后，把该dict转换为关键字参数传进去：

```
>>> extra = {'city': 'Beijing', 'job': 'Engineer'}
>>> person('Jack', 24, city=extra['city'], job=extra['job'])
name: Jack age: 24 other: {'city': 'Beijing', 'job': 'Engineer'}
```

当然，上面复杂的调用可以用简化的写法：

```
>>> extra = {'city': 'Beijing', 'job': 'Engineer'}
>>> person('Jack', 24, **extra)
name: Jack age: 24 other: {'city': 'Beijing', 'job': 'Engineer'}
```

`**extra` 表示把 `extra` 这个 dict 的所有 key-value 用关键字参数传入到函数的 `**kw` 参数，`kw` 将获得一个 dict，注意 `kw` 获得的 dict 是 `extra` 的一份拷贝，对 `kw` 的改动不会影响到函数外的 `extra`。

命名关键字参数

对于关键字参数，函数的调用者可以传入任意不受限制的关键字参数。至于到底传入了哪些，就需要在函数内部通过 `kw` 检查。

仍以 `person()` 函数为例，我们希望检查是否有 `city` 和 `job` 参数：

```
def person(name, age, **kw):
    if 'city' in kw:
        # 有 city 参数
        pass
    if 'job' in kw:
        # 有 job 参数
        pass
    print('name:', name, 'age:', age, 'other:', kw)
```

但是调用者仍可以传入不受限制的关键字参数：

```
>>> person('Jack', 24, city='Beijing', addr='Chaoyang', zipcode=123456)
```

如果要限制关键字参数的名字，就可以用命名关键字参数，例如，只接收 `city` 和 `job` 作为关键字参数。这种方式定义的函数如下：

```
def person(name, age, *, city, job):
    print(name, age, city, job)
```

和关键字参数 `**kw` 不同，命名关键字参数需要一个特殊分隔符 `*`，`*` 后面的参数被视为命名关键字参数。

调用方式如下：

```
>>> person('Jack', 24, city='Beijing', job='Engineer')
Jack 24 Beijing Engineer
```

如果函数定义中已经有了一个可变参数，后面跟着的命名关键字参数就不再需要一个特殊分隔符 `*` 了：

```
def person(name, age, *args, city, job):
    print(name, age, args, city, job)
```

命名关键字参数必须传入参数名，这和位置参数不同。如果没有传入参数名，调用将报错：

```
>>> person('Jack', 24, 'Beijing', 'Engineer')
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: person() takes 2 positional arguments but 4 were given
```

由于调用时缺少参数名 `city` 和 `job`，Python 解释器把这 4 个参数均视为位置参数，但 `person()` 函数仅接受 2 个位置参数。

命名关键字参数可以有缺省值，从而简化调用：

```
def person(name, age, *, city='Beijing', job):
    print(name, age, city, job)
```

由于命名关键字参数 `city` 具有默认值，调用时，可不传入 `city` 参数：

```
>>> person('Jack', 24, job='Engineer')
Jack 24 Beijing Engineer
```

使用命名关键字参数时，要特别注意，如果没有可变参数，就必须加一个 `*` 作为特殊分隔符。如果缺少 `*`，Python 解释器将无法识别位置参数和命名关键字参数：

```
def person(name, age, city, job):
    # 缺少 *, city和job被视为位置参数
    pass
```

参数组合

在 Python 中定义函数，可以用必选参数、默认参数、可变参数、关键字参数和命名关键字参数，这 5 种参数都可以组合使用。但是请注意，参数定义的顺序必须是：必选参数、默认参数、可变参数、命名关键字参数和关键字参数。

比如定义一个函数，包含上述若干种参数：

```
def f1(a, b, c=0, *args, **kw):
    print('a =', a, 'b =', b, 'c =', c, 'args =', args, 'kw =', kw)

def f2(a, b, c=0, *, d, **kw):
    print('a =', a, 'b =', b, 'c =', c, 'd =', d, 'kw =', kw)
```

在函数调用的时候，Python 解释器自动按照参数位置和参数名把对应的参数传进去。

```
>>> f1(1, 2)
a = 1 b = 2 c = 0 args = () kw = {}
>>> f1(1, 2, c=3)
a = 1 b = 2 c = 3 args = () kw = {}
>>> f1(1, 2, 3, 'a', 'b')
a = 1 b = 2 c = 3 args = ('a', 'b') kw = {}
>>> f1(1, 2, 3, 'a', 'b', x=99)
a = 1 b = 2 c = 3 args = ('a', 'b') kw = {'x': 99}
>>> f2(1, 2, d=99, ext=None)
a = 1 b = 2 c = 0 d = 99 kw = {'ext': None}
```

最神奇的是通过一个 tuple 和 dict，你也可以调用上述函数：

```
>>> args = (1, 2, 3, 4)
>>> kw = {'d': 99, 'x': '#'}
>>> f1(*args, **kw)
a = 1 b = 2 c = 3 args = (4,) kw = {'d': 99, 'x': '#'}
>>> args = (1, 2, 3)
>>> kw = {'d': 88, 'x': '#'}
>>> f2(*args, **kw)
a = 1 b = 2 c = 3 d = 88 kw = {'x': '#'}
```

所以，对于任意函数，都可以通过类似 `func(*args, **kw)` 的形式调用它，无论它的参数是如何定义的。

虽然可以组合多达 5 种参数，但不要同时使用太多的组合，否则函数接口的可理解性很差。

练习

以下函数允许计算两个数的乘积，请稍加改造，变成可接收一个或多个数并计算乘积：

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

```

# 原函数
# def product(x, y):
#     return x * y

# 改造后
def product(x, *numbers):
    y = 1
    for y in numbers:
        x = x * y
    return x

# 测试
print('product(5) =', product(5))
print('product(5, 6) =', product(5, 6))
print('product(5, 6, 7) =', product(5, 6, 7))
print('product(5, 6, 7, 9) =', product(5, 6, 7, 9))
if product(5) != 5:
    print('测试失败!')
elif product(5, 6) != 30:
    print('测试失败!')
elif product(5, 6, 7) != 210:
    print('测试失败!')
elif product(5, 6, 7, 9) != 1890:
    print('测试失败!')
else:
    try:
        product()
        print('测试失败!')
    except TypeError:
        print('测试成功!')

```

小结

Python的函数具有非常灵活的参数形态，既可以实现简单的调用，又可以传入非常复杂的参数。

默认参数一定要用不可变对象，如果是可变对象，程序运行时会有逻辑错误！

要注意定义可变参数和关键字参数的语法：

`*args` 是可变参数，`args`接收的是一个tuple；

`**kw` 是关键字参数，`kw`接收的是一个dict。

以及调用函数时如何传入可变参数和关键字参数的语法：

可变参数既可以直接传入：`func(1, 2, 3)`，又可以先组装list或tuple，再通过`*args`传入：`func(*(1, 2, 3))`；

关键字参数既可以直接传入：`func(a=1, b=2)`，又可以先组装dict，再通过`**kw`传入：`func(**{'a': 1, 'b': 2})`。

使用`*args`和`**kw`是Python的习惯写法，当然也可以用其他参数名，但最好使用习惯用法。

命名的关键字参数是为了限制调用者可以传入的参数名，同时可以提供默认值。

定义命名的关键字参数在没有可变参数的情况下不要忘了写分隔符`*`，否则定义的将是位置参数。

参考源码

[var_args.py](#)

[kw_args.py](#)

递归函数

在函数内部，可以调用其他函数。如果一个函数在内部调用自身本身，这个函数就是递归函数。

举个例子，我们来计算阶乘 $n! = 1 \times 2 \times 3 \times \dots \times n$ ，用函数 `fact(n)` 表示，可以看出：

`fact(n) = n! = 1 × 2 × 3 × … × (n-1) × n = (n-1)! × n = fact(n-1) × n`

所以，`fact(n)` 可以表示为 `n × fact(n-1)`，只有 $n=1$ 时需要特殊处理。

于是，`fact(n)` 用递归的方式写出来就是：

```
def fact(n):
    if n==1:
        return 1
    return n * fact(n - 1)
```

上面就是一个递归函数。可以试试：

```
>>> fact(1)
1
>>> fact(5)
120
>>> fact(100)
9332621544394415268169923885626670049071\
5968264381621468592963895217599993229915\
6089414639761565182862536979208272237582\
511852109168640000000000000000000000000000000000
```

如果我们计算 `fact(5)`，可以根据函数定义看到计算过程如下：

```
==> fact(5)
==> 5 * fact(4)
==> 5 * (4 * fact(3))
==> 5 * (4 * (3 * fact(2)))
==> 5 * (4 * (3 * (2 * fact(1))))
==> 5 * (4 * (3 * (2 * 1)))
==> 5 * (4 * (3 * 2))
==> 5 * (4 * 6)
==> 5 * 24
==> 120
```

递归函数的优点是定义简单，逻辑清晰。理论上，所有的递归函数都可以写成循环的方式，但循环的逻辑不如递归清晰。

使用递归函数需要注意防止栈溢出。在计算机中，函数调用是通过栈（stack）这种数据结构实现的，每当进入一个函数调用，栈就会加一层栈帧，每当函数返回，栈就会减一层栈帧。由于栈的大小不是无限的，所以，递归调用的次数过多，会导致栈溢出。可以试试 `fact(1000)`：

```
>>> fact(1000)
Traceback (most recent call last):
File "<stdin>", line 1, in <module>
File "<stdin>", line 4, in fact
...
File "<stdin>", line 4, in fact
RuntimeError: maximum recursion depth exceeded in comparison
```

解决递归调用栈溢出的方法是通过尾递归优化，事实上尾递归和循环的效果是一样的，所以，把循环看成是一种特殊的尾递归函数也是可以的。

尾递归是指，在函数返回的时候，调用自身本身，并且，return语句不能包含表达式。这样，编译器或者解释器就可以把尾递归做优化，使递归本身无论调用多少次，都只占用一个栈帧，不会出现栈溢出的情况。

上面的 fact(n) 函数由于 return n * fact(n - 1) 引入了乘法表达式，所以就不是尾递归了。要改成尾递归方式，需要多一点代码，主要是要把每一步的乘积传入到递归函数中：

```
def fact(n):
    return fact_iter(n, 1)

def fact_iter(num, product):
    if num == 1:
        return product
    return fact_iter(num - 1, num * product)
```

可以看到，return fact_iter(num - 1, num * product) 仅返回递归函数本身，num - 1 和 num * product 在函数调用前就会被计算，不影响函数调用。

fact(5) 对应的 fact_iter(5, 1) 的调用如下：

```
==> fact_iter(5, 1)
==> fact_iter(4, 5)
==> fact_iter(3, 20)
==> fact_iter(2, 60)
==> fact_iter(1, 120)
==> 120
```

尾递归调用时，如果做了优化，栈不会增长，因此，无论多少次调用也不会导致栈溢出。

遗憾的是，大多数编程语言没有针对尾递归做优化，Python解释器也没有做优化，所以，即使把上面的 fact(n) 函数改成尾递归方式，也会导致栈溢出。

小结

使用递归函数的优点是逻辑简单清晰，缺点是过深的调用会导致栈溢出。

针对尾递归优化的语言可以通过尾递归防止栈溢出。尾递归事实上和循环是等价的，没有循环语句的编程语言只能通过尾递归实现循环。

Python标准的解释器没有针对尾递归做优化，任何递归函数都存在栈溢出的问题。

练习

[汉诺塔](#)的移动可以用递归函数非常简单地实现。

请编写 move(n, a, b, c) 函数，它接收参数 n，表示3个柱子A、B、C中第1个柱子A的盘子数量，然后打印出把所有盘子从A借助B移动到C的方法，例如：

```
# -*- coding: utf-8 -*-
def move(n, a, b, c):

    if n == 1:
        print(a, '-->', c)
    else:
        move(n-1, a, c, b)
        move(1, a, b, c)
        move(n-1, b, a, c)
```

```
# 期待输出:  
# 期待输出:  
# A --> C  
# A --> B  
# C --> B  
# A --> C  
# B --> A  
# B --> C  
# A --> C  
move(3, 'A', 'B', 'C')move(3, 'A', 'B', 'C')
```

参考源码

[recur.py](#)

高级特性

掌握了Python的数据类型、语句和函数，基本上就可以编写出很多有用的程序了。

比如构造一个 `1, 3, 5, 7, ..., 99` 的列表，可以通过循环实现：

```
L = []
n = 1
while n <= 99:
    L.append(n)
    n = n + 2
```

取list的前一半的元素，也可以通过循环实现。

但是在Python中，代码不是越多越好，而是越少越好。代码不是越复杂越好，而是越简单越好。

基于这一思想，我们来介绍Python中非常有用的高级特性，1行代码能实现的功能，决不写5行代码。请始终牢记，代码越少，开发效率越高。

切片

取一个list或tuple的部分元素是非常常见的操作。比如，一个list如下：

```
>>> L = ['Michael', 'Sarah', 'Tracy', 'Bob', 'Jack']
```

取前3个元素，应该怎么做？

笨办法：

```
>>> [L[0], L[1], L[2]]
['Michael', 'Sarah', 'Tracy']
```

之所以是笨办法是因为扩展一下，取前N个元素就没辙了。

取前N个元素，也就是索引为0-(N-1)的元素，可以用循环：

```
>>> r = []
>>> n = 3
>>> for i in range(n):
...     r.append(L[i])
...
>>> r
['Michael', 'Sarah', 'Tracy']
```

对这种经常取指定索引范围的操作，用循环十分繁琐，因此，Python提供了切片（Slice）操作符，能大大简化这种操作。

对应上面的问题，取前3个元素，用一行代码就可以完成切片：

```
>>> L[0:3]
['Michael', 'Sarah', 'Tracy']
```

`L[0:3]` 表示，从索引 `0` 开始取，直到索引 `3` 为止，但不包括索引 `3`。即索引 `0`, `1`, `2`，正好是3个元素。

如果第一个索引是 `0`，还可以省略：

```
>>> L[:3]
['Michael', 'Sarah', 'Tracy']
```

也可以从索引1开始，取出2个元素出来：

```
>>> L[1:3]
['Sarah', 'Tracy']
```

类似的，既然Python支持 `L[-1]` 取倒数第一个元素，那么它同样支持倒数切片，试试：

```
>>> L[-2:]
['Bob', 'Jack']
>>> L[-2:-1]
['Bob']
```

记住倒数第一个元素的索引是 `-1`。

切片操作十分有用。我们先创建一个0-99的数列：

```
>>> L = list(range(100))
>>> L
[0, 1, 2, 3, ..., 99]
```

可以通过切片轻松取出某一段数列。比如前10个数：

```
>>> L[:10]
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
```

后10个数：

```
>>> L[-10:]
[90, 91, 92, 93, 94, 95, 96, 97, 98, 99]
```

前11-20个数：

```
>>> L[10:20]
[10, 11, 12, 13, 14, 15, 16, 17, 18, 19]
```

前10个数，每两个取一个：

```
>>> L[:10:2]
[0, 2, 4, 6, 8]
```

所有数，每5个取一个：

```
>>> L[::-5]
[0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95]
```

甚至什么都不写，只写 `[:]` 就可以原样复制一个list：

```
>>> L[:]
[0, 1, 2, 3, ..., 99]
```

tuple也是一种list，唯一区别是tuple不可变。因此，tuple也可以用切片操作，只是操作的结果仍是tuple：

```
>>> (0, 1, 2, 3, 4, 5)[:3]
(0, 1, 2)
```

字符串 `'xxx'` 也可以看成是一种list，每个元素就是一个字符。因此，字符串也可以用切片操作，只是操作结果仍是字符串：

```
>>> 'ABCDEFG'[:3]
'ABC'
>>> 'ABCDEFG'[::-2]
'ACEG'
```

在很多编程语言中，针对字符串提供了很多各种截取函数（例如，`substring`），其实目的就是对字符串切片。Python没有针对字符串的截取函数，只需要切片一个操作就可以完成，非常简单。

练习

利用切片操作，实现一个trim()函数，去除字符串首尾的空格，注意不要调用str的 strip() 方法：

```
# -*- coding: utf-8 -*-
def trim(s):
    while s[:1] == ' ':
        s = s[1:]
    while s[-1:] == ' ':
        s = s[:-1]
    return s

# 测试:
if trim('hello ') != 'hello':
    print('测试失败!')
elif trim(' hello') != 'hello':
    print('测试失败!')
elif trim(' hello ') != 'hello':
    print('测试失败!')
elif trim(' hello world ') != 'hello world':
    print('测试失败!')
elif trim('') != '':
    print('测试失败!')
elif trim(' ') != '':
    print('测试失败!')
else:
    print('测试成功!')
```

小结

有了切片操作，很多地方循环就不再需要了。Python的切片非常灵活，一行代码就可以实现很多行循环才能完成的操作。

参考源码

[do_slice.py](#)

迭代

如果给定一个list或tuple，我们可以通过for循环来遍历这个list或tuple，这种遍历我们称为迭代（Iteration）。

在Python中，迭代是通过for ... in 来完成的，而很多语言比如C语言，迭代list是通过下标完成的，比如Java代码：

```
for (i=0; i<list.length; i++) {
    n = list[i];
}
```

可以看出，Python的for循环抽象程度要高于C的for循环，因为Python的for循环不仅可以用在list或tuple上，还可以作用在其他可迭代对象上。

list这种数据类型虽然有下标，但很多其他数据类型是没有下标的，但是，只要是可迭代对象，无论有无下标，都可以迭代，比如dict就可以迭代：

```
>>> d = {'a': 1, 'b': 2, 'c': 3}
>>> for key in d:
...     print(key)
...
a
c
b
```

因为dict的存储不是按照list的方式顺序排列，所以，迭代出的结果顺序很可能不一样。

默认情况下，dict迭代的是key。如果要迭代value，可以用for value in d.values()，如果要同时迭代key和value，可以用for k, v in d.items()。

由于字符串也是可迭代对象，因此，也可以作用于for循环：

```
>>> for ch in 'ABC':
...     print(ch)
...
A
B
C
```

所以，当我们使用for循环时，只要作用于一个可迭代对象，for循环就可以正常运行，而我们不太关心该对象究竟是list还是其他数据类型。

那么，如何判断一个对象是可迭代对象呢？方法是通过collections模块的Iterable类型判断：

```
>>> from collections import Iterable
>>> isinstance('abc', Iterable) # str是否可迭代
True
>>> isinstance([1,2,3], Iterable) # list是否可迭代
True
>>> isinstance(123, Iterable) # 整数是否可迭代
False
```

最后一个小问题，如果要对list实现类似Java那样的下标循环怎么办？Python内置的enumerate函数可以把一个list变成索引-元素对，这样就可以在for循环中同时迭代索引和元素本身：

```
>>> for i, value in enumerate(['A', 'B', 'C']):
...     print(i, value)
...
```

```

0 A
1 B
2 C

```

上面的 `for` 循环里，同时引用了两个变量，在Python里是很常见的，比如下面的代码：

```

>>> for x, y in [(1, 1), (2, 4), (3, 9)]:
...     print(x, y)
...
1 1
2 4
3 9

```

练习

请使用迭代查找一个list中最小和最大值，并返回一个tuple：

```

# -*- coding: utf-8 -*-
def findMinAndMax(L):
    if L == []:
        return (None, None)
    min = max = L[0]
    for x in L:
        if max <= x:
            max = x
        if min >= x:
            min = x
    return (min, max)

# 测试
if findMinAndMax([]) != (None, None):
    print('测试失败!')
elif findMinAndMax([7]) != (7, 7):
    print('测试失败!')
elif findMinAndMax([7, 1]) != (1, 7):
    print('测试失败!')
elif findMinAndMax([7, 1, 3, 9, 5]) != (1, 9):
    print('测试失败!')
else:
    print('测试成功!')

```

小结

任何可迭代对象都可以作用于 `for` 循环，包括我们自定义的数据类型，只要符合迭代条件，就可以使用 `for` 循环。

参考源码

[do_iter.py](#)

列表生成式

列表生成式即List Comprehensions，是Python内置的非常简单却强大的可以用来创建list的生成式。

举个例子，要生成list [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] 可以用 list(range(1, 11))：

```
>>> list(range(1, 11))
[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
```

但如果要生成 [1x1, 2x2, 3x3, ..., 10x10] 怎么做？方法一是循环：

```
>>> L = []
>>> for x in range(1, 11):
...     L.append(x * x)
...
>>> L
[1, 4, 9, 16, 25, 36, 49, 64, 81, 100]
```

但是循环太繁琐，而列表生成式则可以用一行语句代替循环生成上面的list：

```
>>> [x * x for x in range(1, 11)]
[1, 4, 9, 16, 25, 36, 49, 64, 81, 100]
```

写列表生成式时，把要生成的元素 `x * x` 放到前面，后面跟 `for` 循环，就可以把list创建出来，十分有用，多写几次，很快就可以熟悉这种语法。

`for` 循环后面还可以加上`if`判断，这样我们就可以筛选出仅偶数的平方：

```
>>> [x * x for x in range(1, 11) if x % 2 == 0]
[4, 16, 36, 64, 100]
```

还可以使用两层循环，可以生成全排列：

```
>>> [m + n for m in 'ABC' for n in 'XYZ']
['AX', 'AY', 'AZ', 'BX', 'BY', 'BZ', 'CX', 'CY', 'CZ']
```

三层和三层以上的循环就很少用到了。

运用列表生成式，可以写出非常简洁的代码。例如，列出当前目录下的所有文件和目录名，可以通过一行代码实现：

```
>>> import os # 导入os模块，模块的概念后面讲到
>>> [d for d in os.listdir('.')] # os.listdir可以列出文件和目录
['.emacs.d', '.ssh', '.Trash', 'Adlm', 'Applications', 'Desktop', 'Documents', 'Downloads', 'Library', 'Movies', '']
```

`for` 循环其实可以同时使用两个甚至多个变量，比如 `dict` 的 `items()` 可以同时迭代key和value：

```
>>> d = {'x': 'A', 'y': 'B', 'z': 'C'}
>>> for k, v in d.items():
...     print(k, '=', v)
...
y = B
x = A
z = C
```

因此，列表生成式也可以使用两个变量来生成list：

```
>>> d = {'x': 'A', 'y': 'B', 'z': 'C'}
>>> [k + '=' + v for k, v in d.items()]
['y=B', 'x=A', 'z=C']
```

最后把一个list中所有的字符串变成小写：

```
>>> L = ['Hello', 'World', 'IBM', 'Apple']
>>> [s.lower() for s in L]
['hello', 'world', 'ibm', 'apple']
```

练习

如果list中既包含字符串，又包含整数，由于非字符串类型没有 `lower()` 方法，所以列表生成式会报错：

```
>>> L = ['Hello', 'World', 18, 'Apple', None]
>>> [s.lower() for s in L]
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
    File "<stdin>", line 1, in <listcomp>
AttributeError: 'int' object has no attribute 'lower'
```

使用内建的 `isinstance` 函数可以判断一个变量是不是字符串：

```
>>> x = 'abc'
>>> y = 123
>>> isinstance(x, str)
True
>>> isinstance(y, str)
False
```

请修改列表生成式，通过添加 `if` 语句保证列表生成式能正确地执行：

```
# -*- coding: utf-8 -*-
L1 = ['Hello', 'World', 18, 'Apple', None]

L2 = [s.lower() for s in L1 if isinstance(s, str)]

# 测试:
print(L2)
if L2 == ['hello', 'world', 'apple']:
    print('测试通过!')
else:
    print('测试失败!')
```

小结

运用列表生成式，可以快速生成list，可以通过一个list推导出另一个list，而代码却十分简洁。

参考源码

[do_listcompr.py](#)

生成器

通过列表生成式，我们可以直接创建一个列表。但是，受到内存限制，列表容量肯定是有限的。而且，创建一个包含100万个元素的列表，不仅占用很大的存储空间，如果我们仅仅需要访问前面几个元素，那后面绝大多数元素占用的空间都白白浪费了。

所以，如果列表元素可以按照某种算法推算出来，那我们是否可以在循环的过程中不断推算出后续的元素呢？这样就不必创建完整的list，从而节省大量的空间。在Python中，这种一边循环一边计算的机制，称为生成器：generator。

要创建一个generator，有很多种方法。第一种方法很简单，只要把一个列表生成式的`[]`改成`()`，就创建了一个generator：

```
>>> L = [x * x for x in range(10)]
>>> L
[0, 1, 4, 9, 16, 25, 36, 49, 64, 81]
>>> g = (x * x for x in range(10))
>>> g
<generator object <genexpr> at 0x1022ef630>
```

创建`L`和`g`的区别仅在于最外层的`[]`和`()`，`L`是一个list，而`g`是一个generator。

我们可以直接打印出list的每一个元素，但我们怎么打印出generator的每一个元素呢？

如果要一个一个打印出来，可以通过`next()`函数获得generator的下一个返回值：

```
>>> next(g)
0
>>> next(g)
1
>>> next(g)
4
>>> next(g)
9
>>> next(g)
16
>>> next(g)
25
>>> next(g)
36
>>> next(g)
49
>>> next(g)
64
>>> next(g)
81
>>> next(g)
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
StopIteration
```

我们讲过，generator保存的是算法，每次调用`next(g)`，就计算出`g`的下一个元素的值，直到计算到最后一个元素，没有更多的元素时，抛出`StopIteration`的错误。

当然，上面这种不断调用`next(g)`实在是太变态了，正确的方法是使用`for`循环，因为generator也是可迭代对象：

```
>>> g = (x * x for x in range(10))
>>> for n in g:
...     print(n)
...
0
```

```
1
4
9
16
25
36
49
64
81
```

所以，我们创建了一个generator后，基本上永远不会调用 `next()`，而是通过 `for` 循环来迭代它，并且不需要关心 `StopIteration` 的错误。

generator非常强大。如果推算的算法比较复杂，用类似列表生成式的 `for` 循环无法实现的时候，还可以用函数来实现。

比如，著名的斐波拉契数列（Fibonacci），除第一个和第二个数外，任意一个数都可由前两个数相加得到：

1, 1, 2, 3, 5, 8, 13, 21, 34, ...

斐波拉契数列用列表生成式写不出来，但是，用函数把它打印出来却很容易：

```
def fib(max):
    n, a, b = 0, 0, 1
    while n < max:
        print(b)
        a, b = b, a + b
        n = n + 1
    return 'done'
```

注意，赋值语句

```
a, b = b, a + b
```

相当于：

```
t = (b, a + b) # t是一个tuple
a = t[0]
b = t[1]
```

但不必显式写出临时变量t就可以赋值。

上面的函数可以输出斐波那契数列的前N个数：

```
>>> fib(6)
1
1
2
3
5
8
'done'
```

仔细观察，可以看出，`fib` 函数实际上是定义了斐波拉契数列的推算规则，可以从第一个元素开始，推算出后续任意的元素，这种逻辑其实非常类似generator。

也就是说，上面的函数和generator仅一步之遥。要把 `fib` 函数变成generator，只需要把 `print(b)` 改为 `yield b` 就可以了：

```
def fib(max):
```

```

n, a, b = 0, 0, 1
while n < max:
    yield b
    a, b = b, a + b
    n = n + 1
return 'done'

```

这就是定义generator的另一种方法。如果一个函数定义中包含 `yield` 关键字，那么这个函数就不再是一个普通函数，而是一个generator：

```

>>> f = fib(6)
>>> f
<generator object fib at 0x104feaaa0>

```

这里，最难理解的就是generator和函数的执行流程不一样。函数是顺序执行，遇到 `return` 语句或者最后一行函数语句就返回。而变成generator的函数，在每次调用 `next()` 的时候执行，遇到 `yield` 语句返回，再次执行时从上次返回的 `yield` 语句处继续执行。

举个简单的例子，定义一个generator，依次返回数字1, 3, 5：

```

def odd():
    print('step 1')
    yield 1
    print('step 2')
    yield(3)
    print('step 3')
    yield(5)

```

调用该generator时，首先要生成一个generator对象，然后用 `next()` 函数不断获得下一个返回值：

```

>>> o = odd()
>>> next(o)
step 1
1
>>> next(o)
step 2
3
>>> next(o)
step 3
5
>>> next(o)
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
StopIteration

```

可以看到，`odd` 不是普通函数，而是generator，在执行过程中，遇到 `yield` 就中断，下次又继续执行。执行3次 `yield` 后，已经没有 `yield` 可以执行了，所以，第4次调用 `next(o)` 就报错。

回到 `fib` 的例子，我们在循环过程中不断调用 `yield`，就会不断中断。当然要给循环设置一个条件来退出循环，不然就会产生一个无限数列出来。

同样的，把函数改成generator后，我们基本上从来不会用 `next()` 来获取下一个返回值，而是直接使用 `for` 循环来迭代：

```

>>> for n in fib(6):
...     print(n)
...
1
1
2

```

```
3
5
8
```

但是用 `for` 循环调用 generator 时，发现拿不到 generator 的 `return` 语句的返回值。如果想要拿到返回值，必须捕获 `StopIteration` 错误，返回值包含在 `StopIteration` 的 `value` 中：

```
>>> g = fib(6)
>>> while True:
...     try:
...         x = next(g)
...         print('g:', x)
...     except StopIteration as e:
...         print('Generator return value:', e.value)
...         break
...
g: 1
g: 1
g: 2
g: 3
g: 5
g: 8
Generator return value: done
```

关于如何捕获错误，后面错误处理还会详细讲解。

练习

[杨辉三角](#) 定义如下：

```
1
/
1 1
/ \ / \
1 2 1
/ \ / \ /
1 3 3 1
/ \ / \ / \ \
1 4 6 4 1
/ \ / \ / \ / \
1 5 10 10 5 1
```

把每一行看做一个 list，试写一个 generator，不断输出下一行的 list：

```
# -*- coding: utf-8 -*-
def triangles():
    num = [1]
    while True:
        yield num
        num = [num[i] + num[i-1] for i in range(len(num)) if i != 0]
        num.append(1)
        num.insert(0, 1)
# 期待输出:
# [1]
# [1, 1]
# [1, 2, 1]
# [1, 3, 3, 1]
# [1, 4, 6, 4, 1]
# [1, 5, 10, 10, 5, 1]
# [1, 6, 15, 20, 15, 6, 1]
# [1, 7, 21, 35, 35, 21, 7, 1]
# [1, 8, 28, 56, 70, 56, 28, 8, 1]
```

```

# [1, 9, 36, 84, 126, 126, 84, 36, 9, 1]
n = 0
results = []
for t in triangles():
    print(t)
    results.append(t)
    n = n + 1
    if n == 10:
        break
if results == [
    [1],
    [1, 1],
    [1, 2, 1],
    [1, 3, 3, 1],
    [1, 4, 6, 4, 1],
    [1, 5, 10, 10, 5, 1],
    [1, 6, 15, 20, 15, 6, 1],
    [1, 7, 21, 35, 35, 21, 7, 1],
    [1, 8, 28, 56, 70, 56, 28, 8, 1],
    [1, 9, 36, 84, 126, 126, 84, 36, 9, 1]
]:
    print('测试通过!')
else:
    print('测试失败!')

```

小结

generator是非常强大的工具，在Python中，可以简单地把列表生成式改成generator，也可以通过函数实现复杂逻辑的generator。

要理解generator的工作原理，它是在 `for` 循环的过程中不断计算出下一个元素，并在适当的条件结束 `for` 循环。对于函数改成的generator来说，遇到 `return` 语句或者执行到函数体最后一行语句，就是结束generator的指令，`for` 循环随之结束。

请注意区分普通函数和generator函数，普通函数调用直接返回结果：

```

>>> r = abs(6)
>>> r
6

```

generator函数的“调用”实际返回一个generator对象：

```

>>> g = fib(6)
>>> g
<generator object fib at 0x1022ef948>

```

参考源码

[do_generator.py](#)

迭代器

我们已经知道，可以直接作用于 `for` 循环的数据类型有以下几种：

一类是集合数据类型，如 `list`、`tuple`、`dict`、`set`、`str` 等；

一类是 `generator`，包括生成器和带 `yield` 的 `generator function`。

这些可以直接作用于 `for` 循环的对象统称为可迭代对象： `Iterable`。

可以使用 `isinstance()` 判断一个对象是否是 `Iterable` 对象：

```
>>> from collections import Iterable
>>> isinstance([], Iterable)
True
>>> isinstance({}, Iterable)
True
>>> isinstance('abc', Iterable)
True
>>> isinstance((x for x in range(10)), Iterable)
True
>>> isinstance(100, Iterable)
False
```

而生成器不但可以作用于 `for` 循环，还可以被 `next()` 函数不断调用并返回下一个值，直到最后抛出 `StopIteration` 错误表示无法继续返回下一个值了。

可以被 `next()` 函数调用并不断返回下一个值的对象称为迭代器：`Iterator`。

可以使用 `isinstance()` 判断一个对象是否是 `Iterator` 对象：

```
>>> from collections import Iterator
>>> isinstance((x for x in range(10)), Iterator)
True
>>> isinstance([], Iterator)
False
>>> isinstance({}, Iterator)
False
>>> isinstance('abc', Iterator)
False
```

生成器都是 `Iterator` 对象，但 `list`、`dict`、`str` 虽然是 `Iterable`，却不是 `Iterator`。

把 `list`、`dict`、`str` 等 `Iterable` 变成 `Iterator` 可以使用 `iter()` 函数：

```
>>> isinstance(iter([]), Iterator)
True
>>> isinstance(iter('abc'), Iterator)
True
```

你可能会问，为什么 `list`、`dict`、`str` 等数据类型不是 `Iterator`？

这是因为 Python 的 `Iterator` 对象表示的是一个数据流，`Iterator` 对象可以被 `next()` 函数调用并不断返回下一个数据，直到没有数据时抛出 `StopIteration` 错误。可以把这个数据流看做是一个有序序列，但我们却不能提前知道序列的长度，只能不断通过 `next()` 函数实现按需计算下一个数据，所以 `Iterator` 的计算是惰性的，只有在需要返回下一个数据时它才会计算。

`Iterator` 甚至可以表示一个无限大的数据流，例如全体自然数。而使用 `list` 是永远不可能存储全体自然数的。

小结

凡是可作用于 `for` 循环的对象都是 `Iterable` 类型；

凡是可作用于 `next()` 函数的对象都是 `Iterator` 类型，它们表示一个惰性计算的序列；

集合数据类型如 `list`、`dict`、`str` 等是 `Iterable` 但不是 `Iterator`，不过可以通过 `iter()` 函数获得一个 `Iterator` 对象。

Python 的 `for` 循环本质上就是通过不断调用 `next()` 函数实现的，例如：

```
for x in [1, 2, 3, 4, 5]:  
    pass
```

实际上完全等价于：

```
# 首先获得Iterator对象:  
it = iter([1, 2, 3, 4, 5])  
# 循环:  
while True:  
    try:  
        # 获得下一个值:  
        x = next(it)  
    except StopIteration:  
        # 遇到StopIteration就退出循环  
        break
```

参考源码

[do_iter.py](#)

函数式编程

模块

面向对象编程

面向对象高级编程

错误、调试和测试

进程和线程

正则表达式

字符串是编程时涉及到的最多的一种数据结构，对字符串进行操作的需求几乎无处不在。比如判断一个字符串是否是合法的Email地址，虽然可以编程提取 @ 前后的子串，再分别判断是否是单词和域名，但这样做不但麻烦，而且代码难以复用。

正则表达式是一种用来匹配字符串的强有力的武器。它的设计思想是用一种描述性的语言来给字符串定义一个规则，凡是符合规则的字符串，我们就认为它“匹配”了，否则，该字符串就是不合法的。

所以我们判断一个字符串是否是合法的Email的方法是：

1. 创建一个匹配Email的正则表达式；
2. 用该正则表达式去匹配用户的输入来判断是否合法。

因为正则表达式也是用字符串表示的，所以，我们要首先了解如何用字符来描述字符。

在正则表达式中，如果直接给出字符，就是精确匹配。用 \d 可以匹配一个数字，\w 可以匹配一个字母或数字，所以：

- '\d' 可以匹配 '007'，但无法匹配 '00A'；
- '\d\d\d' 可以匹配 '010'；
- '\w\w\d' 可以匹配 'py3'；
- 可以匹配任意字符，所以：
- 'py.' 可以匹配 'pyc'、'pyo'、'py!' 等等。

要匹配变长的字符，在正则表达式中，用 * 表示任意个字符（包括0个），用 + 表示至少一个字符，用 ? 表示0个或1个字符，用 {n} 表示n个字符，用 {n,m} 表示n-m个字符：

来看一个复杂的例子：\d{3}\s+\d{3,8}。

我们来从左到右解读一下：

1. \d{3} 表示匹配3个数字，例如 '010'；
2. \s 可以匹配一个空格（也包括Tab等空白符），所以 \s+ 表示至少有一个空格，例如匹配 ' '，' ' 等；
3. \d{3,8} 表示3-8个数字，例如 '1234567'。

综合起来，上面的正则表达式可以匹配以任意个空格隔开的带区号的电话号码。

如果要匹配 '010-12345' 这样的号码呢？由于 '-' 是特殊字符，在正则表达式中，要用 '\\' 转义，所以，上面的正则是 \d{3}\-\d{3,8}。

但是，仍然无法匹配 '010 - 12345'，因为带有空格。所以我们需要更复杂的匹配方式。

进阶

要做更精确地匹配，可以用 [] 表示范围，比如：

- [0-9a-zA-Z_] 可以匹配一个数字、字母或者下划线；
- [0-9a-zA-Z_]+ 可以匹配至少由一个数字、字母或者下划线组成的字符串，比如 'a100'，'0_Z'，'Py3000' 等等；

- `[a-zA-Z_] [0-9a-zA-Z_]*` 可以匹配由字母或下划线开头，后接任意个由一个数字、字母或者下划线组成的字符串，也就是Python合法的变量；

• `[a-zA-Z_] [0-9a-zA-Z_] {0, 19}` 更精确地限制了变量的长度是1-20个字符（前面1个字符+后面最多19个字符）。

`A|B` 可以匹配A或B，所以 `(P|p)ython` 可以匹配 `'Python'` 或者 `'python'`。

`^` 表示行的开头，`\d` 表示必须以数字开头。

`$` 表示行的结束，`\d$` 表示必须以数字结束。

你可能注意到了，`py` 也可以匹配 `'python'`，但是加上 `^py$` 就变成了整行匹配，就只能匹配 `'py'` 了。

re模块

有了准备知识，我们就可以在Python中使用正则表达式了。Python提供 `re` 模块，包含所有正则表达式的功能。由于Python的字符串本身也用 \ 转义，所以要特别注意：

```
s = 'ABC\\-001' # Python的字符串
# 对应的正则表达式字符串变成：
# 'ABC\~-001'
```

因此我们强烈建议使用Python的 `r` 前缀，就不用考虑转义的问题了：

```
s = r'ABC\~-001' # Python的字符串
# 对应的正则表达式字符串不变：
# 'ABC\~-001'
```

先看看如何判断正则表达式是否匹配：

```
>>> import re
>>> re.match(r'^\d{3}~\d{3,8}$', '010-12345')
<sre.SRE_Match object; span=(0, 9), match='010-12345'>
>>> re.match(r'^\d{3}~\d{3,8}$', '010 12345')
>>>
```

`match()` 方法判断是否匹配，如果匹配成功，返回一个 `Match` 对象，否则返回 `None`。常见的判断方法就是：

```
test = '用户输入的字符串'
if re.match(r'正则表达式', test):
    print('ok')
else:
    print('failed')
```

切分字符串

用正则表达式切分字符串比用固定的字符更灵活，请看正常的切分代码：

```
>>> 'a b c'.split(' ')
['a', 'b', '', '', 'c']
```

嗯，无法识别连续的空格，用正则表达式试试：

```
>>> re.split(r'\s+', 'a b c')
['a', 'b', 'c']
```

无论多少个空格都可以正常分割。加入 , 试试：

```
>>> re.split(r'[\s\,]+', 'a,b, c d')
['a', 'b', 'c', 'd']
```

再加入 ; 试试：

```
>>> re.split(r'[\s\,;\]+', 'a,b;; c d')
['a', 'b', 'c', 'd']
```

如果用户输入了一组标签，下次记得用正则表达式来把不规范的输入转化成正确的数组。

分组

除了简单地判断是否匹配之外，正则表达式还有提取子串的强大功能。用 () 表示的就是要提取的分组（Group）。比如：

`^\(\d{3}\)-(\d{3,8})$` 分别定义了两个组，可以直接从匹配的字符串中提取出区号和本地号码：

```
>>> m = re.match(r'^(\d{3})-(\d{3,8})$', '010-12345')
>>> m
<sre.SRE_Match object; span=(0, 9), match='010-12345'>
>>> m.group(0)
'010-12345'
>>> m.group(1)
'010'
>>> m.group(2)
'12345'
```

如果正则表达式中定义了组，就可以在 Match 对象上用 group() 方法提取出子串来。

注意到 group(0) 永远是原始字符串，group(1)、group(2) 表示第1、2、.....个子串。

提取子串非常有用。来看一个更凶残的例子：

```
>>> t = '19:05:30'
>>> m = re.match(r'^(\d{2}):(\d{2}):(\d{2})', t)
>>> m.groups()
('19', '05', '30')
```

这个正则表达式可以直接识别合法的时间。但是有些时候，用正则表达式也无法做到完全验证，比如识别日期：

```
'^(\d{4})-(\d{2})-(\d{2})$'
```

对于 '2-30'，'4-31' 这样的非法日期，用正则还是识别不了，或者说写出来非常困难，这时就需要程序配合识别了。

贪婪匹配

最后需要特别指出的是，正则匹配默认是贪婪匹配，也就是匹配尽可能多的字符。举例如下，匹配出数字后面的 0：

```
>>> re.match(r'^(\d*)(0*)$', '102300').groups()
('102300', '')
```

由于 \d+ 采用贪婪匹配，直接把后面的 0 全部匹配了，结果 0* 只能匹配空字符串了。

必须让 `\d+` 采用非贪婪匹配（也就是尽可能少匹配），才能把后面的 `0` 匹配出来，加个 `?` 就可以让 `\d+` 采用非贪婪匹配：

```
>>> re.match(r'^(\d+?)(0*)$', '102300').groups()
('1023', '00')
```

编译

当我们在Python中使用正则表达式时，`re`模块内部会干两件事情：

1. 编译正则表达式，如果正则表达式的字符串本身不合法，会报错；
2. 用编译后的正则表达式去匹配字符串。

如果一个正则表达式要重复使用几千次，出于效率的考虑，我们可以预编译该正则表达式，接下来重复使用时就不需要编译这个步骤了，直接匹配：

```
>>> import re
# 编译:
>>> re_telephone = re.compile(r'^(\\d{3})-(\\d{3,8})$')
# 使用:
>>> re_telephone.match('010-12345').groups()
('010', '12345')
>>> re_telephone.match('010-8086').groups()
('010', '8086')
```

编译后生成Regular Expression对象，由于该对象自己包含了正则表达式，所以调用对应的方法时不用给出正则字符串。

小结

正则表达式非常强大，要在短短的一节里讲完是不可能的。要讲清楚正则的所有内容，可以写一本厚厚的书了。如果你经常遇到正则表达式的问题，你可能需要一本正则表达式的参考书。

练习

请尝试写一个验证Email地址的正则表达式。版本一应该可以验证出类似的Email：

- `someone@gmail.com`
- `bill.gates@microsoft.com`

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

em=re.compile(r'[\w\.\.]*@\w*\.\com')
if em.match(addr):
    return True
else:
    return False

# 测试:
assert is_valid_email('someone@gmail.com')
assert is_valid_email('bill.gates@microsoft.com')
assert not is_valid_email('bob@example.com')
assert not is_valid_email('mr-bob@example.com')
print('ok')
```

版本二可以提取出带名字的Email地址：

- <Tom Paris>tom@voyager.org => Tom Paris
- bob@example.com => bob

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

# 测试:
assert name_of_email('<Tom Paris> tom@voyager.org') == 'Tom Paris'
assert name_of_email('tom@voyager.org') == 'tom'
print('ok')
```

参考源码

[regex.py](#)

常用内建模块

常用第三方模块

virtualenv

在开发Python应用程序的时候，系统安装的Python3只有一个版本：3.4。所有第三方的包都会被 `pip` 安装到Python3的 `site-packages` 目录下。

如果我们要同时开发多个应用程序，那这些应用程序都会共用一个Python，就是安装在系统的Python 3。如果应用A需要jinja 2.7，而应用B需要jinja 2.6怎么办？

这种情况下，每个应用可能需要各自拥有一套“独立”的Python运行环境。`virtualenv`就是用来为一个应用创建一套“隔离”的Python运行环境。

首先，我们用 `pip` 安装`virtualenv`：

```
$ pip3 install virtualenv
```

然后，假定我们要开发一个新的项目，需要一套独立的Python运行环境，可以这么做：

第一步，创建目录：

```
Mac:~ michael$ mkdir myproject
Mac:~ michael$ cd myproject/
Mac:myproject michael$
```

第二步，创建一个独立的Python运行环境，命名为 `venv`：

```
Mac:myproject michael$ virtualenv --no-site-packages venv
Using base prefix '/usr/local/.../Python.framework/Versions/3.4'
New python executable in venv/bin/python3.4
Also creating executable in venv/bin/python
Installing setuptools, pip, wheel...done.
```

命令 `virtualenv` 就可以创建一个独立的Python运行环境，我们还加上了参数 `--no-site-packages`，这样，已经安装到系统Python环境中的所有第三方包都不会复制过来，这样，我们就得到了一个不带任何第三方包的“干净”的Python运行环境。

新建的Python环境被放到当前目录下的 `venv` 目录。有了 `venv` 这个Python环境，可以用 `source` 进入该环境：

```
Mac:myproject michael$ source venv/bin/activate
(venv)Mac:myproject michael$
```

注意到命令提示符变了，有个 `(venv)` 前缀，表示当前环境是一个名为 `venv` 的Python环境。

下面正常安装各种第三方包，并运行 `python` 命令：

```
(venv)Mac:myproject michael$ pip install jinja2
...
Successfully installed jinja2-2.7.3 markupsafe-0.23
(venv)Mac:myproject michael$ python myapp.py
...
```

在 `venv` 环境下，用 `pip` 安装的包都被安装到 `venv` 这个环境下，系统Python环境不受任何影响。也就是说，`venv` 环境是专门针对 `myproject` 这个应用创建的。

退出当前的 `venv` 环境，使用 `deactivate` 命令：

```
(venv)Mac:myproject michael$ deactivate  
Mac:myproject michael$
```

此时就回到了正常的环境，现在 `pip` 或 `python` 均是在系统Python环境下执行。

完全可以针对每个应用创建独立的Python运行环境，这样就可以对每个应用的Python环境进行隔离。

`virtualenv`是如何创建“独立”的Python运行环境的呢？原理很简单，就是把系统Python复制一份到`virtualenv`的环境，用命令 `source venv/bin/activate` 进入一个`virtualenv`环境时，`virtualenv`会修改相关环境变量，让命令 `python` 和 `pip` 均指向当前的`virtualenv`环境。

小结

`virtualenv`为应用提供了隔离的Python运行环境，解决了不同应用间多版本的冲突问题。

图形界面

Python 支持多种图形界面的第三方库，包括：

- Tk
- wxWidgets
- Qt
- GTK

等等。

但是 Python 自带的库是支持 Tk 的 Tkinter，使用 Tkinter，无需安装任何包，就可以直接使用。本章简单介绍如何使用 Tkinter 进行 GUI 编程。

Tkinter

我们来梳理一下概念：

我们编写的 Python 代码会调用内置的 Tkinter，Tkinter 封装了访问 Tk 的接口；

Tk 是一个图形库，支持多个操作系统，使用 Tcl 语言开发；

Tk 会调用操作系统提供的本地 GUI 接口，完成最终的 GUI。

所以，我们的代码只需要调用 Tkinter 提供的接口就可以了。

第一个 GUI 程序

使用 Tkinter 十分简单，我们来编写一个 GUI 版本的“Hello, world!”。

第一步是导入 Tkinter 包的所有内容：

```
from tkinter import *
```

第二步是从 `Frame` 派生一个 `Application` 类，这是所有 Widget 的父容器：

```
class Application(Frame):
    def __init__(self, master=None):
        Frame.__init__(self, master)
        self.pack()
        self.createWidgets()

    def createWidgets(self):
        self.helloLabel = Label(self, text='Hello, world!')
        self.helloLabel.pack()
        self.quitButton = Button(self, text='Quit', command=self.quit)
        self.quitButton.pack()
```

在 GUI 中，每个 `Button`、`Label`、输入框等，都是一个 Widget。`Frame` 则是可以容纳其他 Widget 的 Widget，所有的 Widget 组合起来就是一棵树。

`pack()` 方法把 Widget 加入到父容器中，并实现布局。`pack()` 是最简单的布局，`grid()` 可以实现更复杂的布局。

在 `createWidgets()` 方法中，我们创建一个 `Label` 和一个 `Button`，当 `Button` 被点击时，触发 `self.quit()` 使程序退出。

第三步，实例化 `Application`，并启动消息循环：

```
app = Application()
# 设置窗口标题:
app.master.title('Hello World')
# 主消息循环:
app.mainloop()
```

GUI程序的主线程负责监听来自操作系统的消息，并依次处理每一条消息。因此，如果消息处理非常耗时，就需要在新线程中处理。

运行这个GUI程序，可以看到下面的窗口：



点击“Quit”按钮或者窗口的“x”结束程序。

输入文本

我们再对这个GUI程序改进一下，加入一个文本框，让用户可以输入文本，然后点按钮后，弹出消息对话框。

```
from tkinter import *
import tkinter.messagebox as messagebox

class Application(Frame):
    def __init__(self, master=None):
        Frame.__init__(self, master)
        self.pack()
        self.createWidgets()

    def createWidgets(self):
        self.nameInput = Entry(self)
        self.nameInput.pack()
        self.alertButton = Button(self, text='Hello', command=self.hello)
        self.alertButton.pack()

    def hello(self):
        name = self.nameInput.get() or 'world'
        messagebox.showinfo('Message', 'Hello, %s' % name)

app = Application()
# 设置窗口标题:
app.master.title('Hello World')
# 主消息循环:
app.mainloop()
```

当用户点击按钮时，触发 `hello()`，通过 `self.nameInput.get()` 获得用户输入的文本后，使用 `tkMessageBox.showinfo()` 可以弹出消息对话框。

程序运行结果如下：



小结

Python内置的Tkinter可以满足基本的GUI程序的要求，如果是非常复杂的GUI程序，建议用操作系统原生支持的语言和库来编写。

参考源码

[hello_gui.py](#)

网络编程

电子邮件

访问数据库

Web开发

异步IO

实战

FAQ

常见问题

本节列出常见的一些问题。

如何获取当前路径

当前路径可以用 `..` 表示，再用 `os.path.abspath()` 将其转换为绝对路径：

```
# -*- coding:utf-8 -*-
# test.py

import os

print(os.path.abspath('..'))
```

运行结果：

```
$ python3 test.py
/Users/michael/workspace/testing
```

如何获取当前模块的文件名

可以通过特殊变量 `__file__` 获取：

```
# -*- coding:utf-8 -*-
# test.py

print(__file__)
```

输出：

```
$ python3 test.py
test.py
```

如何获取命令行参数

可以通过 `sys` 模块的 `argv` 获取：

```
# -*- coding:utf-8 -*-
# test.py

import sys

print(sys.argv)
```

输出：

```
$ python3 test.py -a -s "Hello world"
['test.py', '-a', '-s', 'Hello world']
```

`argv` 的第一个元素永远是命令行执行的 `.py` 文件名。

如何获取当前Python命令的可执行文件路径

sys 模块的 executable 变量就是Python命令可执行文件的路径：

```
# -*- coding:utf-8 -*-
# test.py

import sys

print(sys.executable)
```

在Mac下的结果：

```
$ python3 test.py
/usr/local/opt/python3/bin/python3.4
```

期末总结

终于到了期末总结的时刻了！

经过一段时间的学习，相信你对Python已经初步掌握。一开始，可能觉得Python上手很容易，可是越往后学，会越困难，有的时候，发现理解不了代码，这时，不妨停下来思考一下，先把概念搞清楚，代码自然就明白了。

Python非常适合初学者用来进入计算机编程领域。Python属于非常高级的语言，掌握了这门高级语言，就对计算机编程的核心思想——抽象有了初步理解。如果希望继续深入学习计算机编程，可以学习Java、C、JavaScript、Lisp等不同类型的语言，只有多掌握不同领域的语言，有比较才更有收获。

