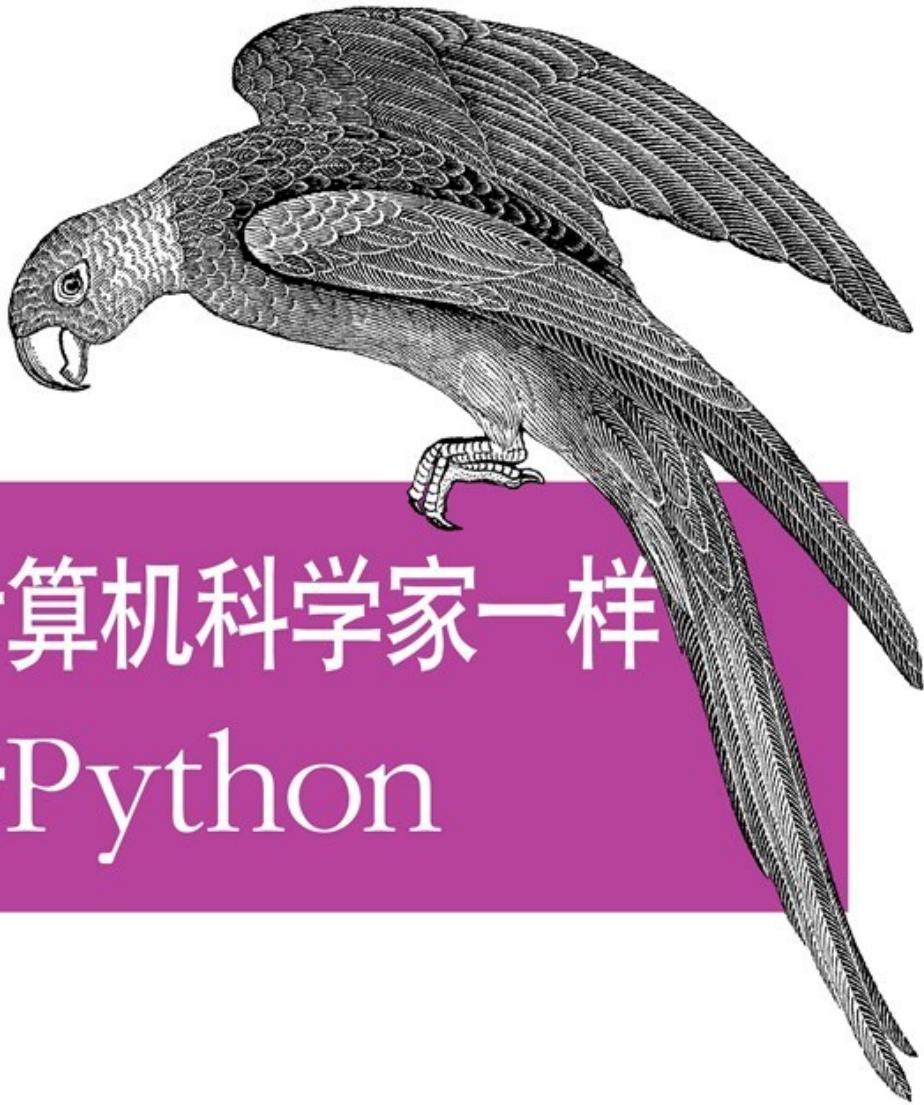


---

*Think Python*



# 像计算机科学家一样 思考Python

O'REILLY®

[美] Allen B. Downey 著  
赵普明 译

 人民邮电出版社  
POSTS & TELECOM PRESS

# 目 录

[版权信息](#)

[版权声明](#)

[内容提要](#)

[前言](#)

[本书的奇特历史](#)

[关于作者](#)

[关于封面](#)

[译后记](#)

[关于译者](#)

[致谢](#)

[贡献者列表](#)

[O'Reilly Media, Inc. 介绍](#)

[第1章 程序之道](#)

[1.1 Python编程语言](#)

[1.2 什么是程序](#)

[1.3 什么是调试](#)

[1.4 语法错误](#)

[1.5 运行时错误](#)

[1.6 语义错误](#)

[1.7 实验型调试](#)

[1.8 形式语言和自然语言](#)

[1.9 第一个程序](#)

[1.10 调试](#)

[1.11 术语表](#)

[1.12 练习](#)

[第2章 变量、表达式和语句](#)

[2.1 值和类型](#)

[2.2 变量](#)

[2.3 变量名称和关键字](#)

[2.4 操作符和操作对象](#)

[2.5 表达式和语句](#)

[2.6 交互模式和脚本模式](#)

[2.7 操作顺序](#)

[2.8 字符串操作](#)

[2.9 注释](#)

[2.10 调试](#)

[2.11 术语表](#)

[2.12 练习](#)

[第3章 函数](#)

[3.1 函数调用](#)

[3.2 类型转换函数](#)

[3.3 数学函数](#)

[3.4 组合](#)

[3.5 添加新函数](#)

[3.6 定义和使用](#)

[3.7 执行流程](#)

[3.8 形参和实参①](#)

[3.9 变量和形参是局部的](#)

[3.10 栈图](#)

[3.11 有返回值函数和无返回值函数](#)

[3.12 为什么要有函数](#)

[3.13 使用from导入模块](#)

[3.14 调试](#)

[3.15 术语表](#)

[3.16 练习](#)

[第4章 案例研究：接口设计](#)

[4.1 乌龟世界](#)

[4.2 简单重复](#)

[4.3 练习](#)

[4.4 封装](#)

[4.5 泛化](#)

[4.6 接口设计](#)

[4.7 重构](#)

[4.8 一个开发计划](#)

[4.9 文档字符串](#)

[4.10 调试](#)

[4.11 术语表](#)

[4.12 练习](#)

[第5章 条件和递归](#)

[5.1 求模操作符](#)

- [5.2 布尔表达式](#)
- [5.3 逻辑操作符](#)
- [5.4 条件执行](#)
- [5.5 选择执行](#)
- [5.6 条件链](#)
- [5.7 嵌套条件](#)
- [5.8 递归](#)
- [5.9 递归函数的栈图](#)
- [5.10 无限递归](#)
- [5.11 键盘输入](#)
- [5.12 调试](#)
- [5.13 术语表](#)
- [5.14 练习](#)
- [第6章 有返回函数](#)
- [6.1 返回值](#)
- [6.2 增量开发](#)
- [6.3 组合](#)
- [6.4 布尔函数](#)
- [6.5 再谈递归](#)
- [6.6 坚持信念](#)
- [6.7 另一个示例](#)
- [6.8 检查类型](#)
- [6.9 调试](#)
- [6.10 术语表](#)
- [6.11 练习](#)
- [第7章 迭代](#)
- [7.1 多重赋值](#)
- [7.2 更新变量](#)
- [7.3 while语句](#)
- [7.4 break语句](#)
- [7.5 平方根](#)
- [7.6 算法](#)
- [7.7 调试](#)
- [7.8 术语表](#)
- [7.9 练习](#)
- [第8章 字符串](#)
- [8.1 字符串是一个序列](#)

[8.2 len](#)

[8.3 使用for循环进行遍历](#)

[8.4 字符串切片](#)

[8.5 字符串是不可变的](#)

[8.6 搜索](#)

[8.7 循环和计数](#)

[8.8 字符串方法](#)

[8.9 操作符in](#)

[8.10 字符串比较](#)

[8.11 调试](#)

[8.12 术语表](#)

[8.13 练习](#)

[第9章 案例分析：文字游戏](#)

[9.1 读取单词列表](#)

[9.2 练习](#)

[9.3 搜索](#)

[9.4 使用下标循环](#)

[9.5 调试](#)

[9.6 术语表](#)

[9.7 练习](#)

[第10章 列表](#)

[10.1 列表是一个序列](#)

[10.2 列表是可变的](#)

[10.3 遍历一个列表](#)

[10.4 列表操作](#)

[10.5 列表切片](#)

[10.6 列表方法](#)

[10.7 映射、过滤和化简](#)

[10.8 删除元素](#)

[10.9 列表和字符串](#)

[10.10 对象和值](#)

[10.11 别名](#)

[10.12 列表参数](#)

[10.13 调试](#)

[10.14 术语表](#)

[10.15 练习](#)

[第11章 字典](#)

[11.1 使用字典作为计数器集合](#)

[11.2 循环和字典](#)

[11.3 反向查找](#)

[11.4 字典和列表](#)

[11.5 备忘](#)

[11.6 全局变量](#)

[11.7 长整数](#)

[11.8 调试](#)

[11.9 术语表](#)

[11.10 练习](#)

[第12章 元组](#)

[12.1 元组是不可变的](#)

[12.2 元组赋值](#)

[12.3 作为返回值的元组](#)

[12.4 可变长参数元组](#)

[12.5 列表和元组](#)

[12.6 字典和元组](#)

[12.7 比较元组](#)

[12.8 序列的序列](#)

[12.9 调试](#)

[12.10 术语表](#)

[12.11 练习](#)

[第13章 案例研究：选择数据结构](#)

[13.1 单词频率分析](#)

[13.2 随机数](#)

[13.3 单词直方图](#)

[13.4 最常用的单词](#)

[13.5 可选形参](#)

[13.6 字典减法](#)

[13.7 随机单词](#)

[13.8 马尔可夫分析](#)

[13.9 数据结构](#)

[13.10 调试](#)

[13.11 术语表](#)

[13.12 练习](#)

[第14章 文件](#)

[14.1 持久化](#)

[14.2 读和写](#)

[14.3 格式操作符](#)

[14.4 文件名和路径](#)

[14.5 捕获异常](#)

[14.6 数据库](#)

[14.7 封存](#)

[14.8 管道](#)

[14.9 编写模块](#)

[14.10 调试](#)

[14.11 术语表](#)

[14.12 练习](#)

[第15章 类和对象](#)

[15.1 用户定义类型](#)

[15.2 属性](#)

[15.3 矩形](#)

[15.4 作为返回值的实例](#)

[15.5 对象是可变的](#)

[15.6 复制](#)

[15.7 调试](#)

[15.8 术语表](#)

[15.9 练习](#)

[第16章 类和函数](#)

[16.1 时间](#)

[16.2 纯函数](#)

[16.3 修改器](#)

[16.4 原型和计划](#)

[16.5 调试](#)

[16.6 术语表](#)

[16.7 练习](#)

[第17章 类和方法](#)

[17.1 面向对象特性](#)

[17.2 打印对象](#)

[17.3 另一个示例](#)

[17.4 一个更复杂的示例](#)

[17.5 init方法](#)

[17.6 str 方法](#)

[17.7 操作符重载](#)

[17.8 基于类型的分发](#)

[17.9 多态](#)

[17.10 调试](#)

[17.11 接口和实现](#)

[17.12 术语表](#)

[17.13 练习](#)

[第18章 继承](#)

[18.1 卡片对象](#)

[18.2 类属性](#)

[18.3 对比卡牌](#)

[18.4 牌组](#)

[18.5 打印牌组](#)

[18.6 添加、删除、洗牌和排序](#)

[18.7 继承](#)

[18.8 类图](#)

[18.9 调试](#)

[18.10 数据封装](#)

[18.11 术语表](#)

[18.12 练习](#)

[第19章 案例研究：Tkinter](#)

[19.1 GUI](#)

[19.2 按钮和回调](#)

[19.3 画布部件](#)

[19.4 坐标序列](#)

[19.5 更多部件](#)

[19.6 包装部件](#)

[19.7 菜单与Callable](#)

[19.8 绑定](#)

[19.9 调试](#)

[19.10 术语表](#)

[19.11 练习](#)

[附录A 调试](#)

[A.1 语法错误](#)

[我一直进行修改，但没有什么区别](#)

[A.2 运行时错误](#)

[A.2.1 我的程序什么都不做](#)

[A.2.2 我的程序卡死了](#)

- [A. 2. 3 无限循环](#)
- [A. 2. 4 无限递归](#)
- [A. 2. 5 执行流程](#)
- [A. 2. 6 当我运行程序，会得到一个异常](#)
- [A. 2. 7 我添加了太多print语句，被输出淹没了](#)
- [A. 3 语义错误](#)
  - [A. 3. 1 我的程序运行不正确](#)
  - [A. 3. 2 我有一个巨大而复杂的表达式，而它和我预料的不同](#)
  - [A. 3. 3 我有个函数或方法返回值和预期不同](#)
  - [A. 3. 4 我真的真的卡住了，我需要帮助](#)
  - [A. 3. 5 不行，我真的需要帮助](#)
- [附录B 算法分析](#)
  - [B. 1 量级](#)
  - [B. 2 Python基本操作的分析](#)
  - [B. 3 搜索算法的分析](#)
  - [B. 4 散列表](#)
- [附录C Lumpy](#)
  - [C. 1 状态图](#)
  - [C. 2 栈图](#)
  - [C. 3 对象图](#)
  - [C. 4 函数和类对象](#)
  - [C. 5 类图](#)
- [欢迎来到异步社区！](#)
- [异步社区的来历](#)
- [社区里都有什么？](#)
- [购买图书](#)
- [下载资源](#)
- [与作译者互动](#)
- [灵活优惠的购书](#)
- [纸电图书组合购买](#)
- [社区里还可以做什么？](#)
- [提交勘误](#)
- [写作](#)
- [会议活动早知道](#)
- [加入异步](#)

# 版权信息

书名：像计算机科学家一样思考Python

ISBN：978-7-115-32092-6

本书由人民邮电出版社发行数字版。版权所有，侵权必究。

---

您购买的人民邮电出版社电子书仅供您个人使用，未经授权，不得以任何方式复制和传播本书内容。

我们愿意相信读者具有这样的良知和觉悟，与我们共同保护知识产权。

如果购买者有侵权行为，我们可能对该用户实施包括但不限于关闭该帐号等维权措施，并可能追究法律责任。

---

• 著 [美] Allen B. Downey

译 赵普明

责任编辑 杨海玲

• 人民邮电出版社出版发行 北京市丰台区成寿寺路11号

邮编 100164 电子邮件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

• 读者服务热线：(010)81055410

反盗版热线：(010)81055315

# 版权声明

Copyright ©2012 by O'Reilly Media, Inc.

Simplified Chinese Edition, jointly published by  
O'Reilly Media, Inc. and Posts & Telecom Press, 2013.  
Authorized translation of the English edition, 2013 O'Reilly  
Media, Inc., the owner of all rights to publish and sell the  
same.

All rights reserved including the rights of reproduction  
in whole or in part in any form.

本书中文简体版由O'Reilly Media, Inc. 授权人民邮电出版社出  
版。未经出版者书面许可，对本书的任何部分不得以任何方式复制  
或抄袭。

版权所有，侵权必究。

# 内容提要

本书按照培养读者像计算机科学家一样的思维方式的思路来教授Python语言编程。全书贯穿的主体是如何思考、设计、开发的方法，而具体的编程语言，只是提供一个具体场景方便介绍的媒介。它并不是一本介绍语言的书，而是一本介绍编程思想的书。和其他编程设计语言书籍不同，它不拘泥于语言细节，而是尝试从初学者的角度出发，用生动的示例和丰富的练习来引导读者渐入佳境。

作者从最基本的编程概念开始讲起，包括语言的语法和语义，而且每个编程概念都有清晰的定义，引领读者循序渐进地学习变量、表达式、语句、函数和数据结构。此外，书中还探讨了如何处理文件和数据库，如何理解对象、方法和面向对象编程，如何使用调试技巧来修正语法、运行时和语义错误。每一章都配有术语表和练习题，方便读者巩固所学的知识和技巧。此外，每一章都抽出一节来讲解如何调试程序。作者针对每章中所专注的语言特性，或者相关的开发问题，总结了调试的方方面面。可以说这是一种非常有益的创新，让初学编程的读者少走很多弯路。

全书共19章和3个附录，详细介绍了Python语言编程的方方面面。这是一本实用的学习指南，适合没有Python编程经验的程序员阅读，也适合高中或大学的学生、Python爱好者及需要了解编程基础的人阅读。对于第一次接触程序设计的人来说，是一本不可多得的佳作。

# 前言

## 本书的奇特历史

1999年，我正在为一门Java的编程入门课程备课。这门课我已经教过3个学期，感到有些灰心。课程的不及格率太高，即使是那些及格的学生，也只获得了很低的成就。

我发现问题之一是教材。它们太厚，有太多冗余的细节，而针对编程技巧的高阶的指导却很不足。并且学生们都受着“陷阱效应”的苦恼：开头时很容易，也能循序渐进，但接着在第5章左右，整个地板就突然陷落了。新资讯来得太多、来得太快，以至于我必须花费一学期剩下的全部时间来帮助他们拾回丢失的片段。

开课前两周，我决定自己来编写教材。我的目标有以下几个。

- 尽量简短。学生读10页书，比不读50页书要好。
- 注意词汇。我尝试尽量少用术语，并在第一次使用它们时做好定义。
- 循序渐进。为了避免陷阱效应，我抽出了最困难的课题，并把它们划分成更细的学习步骤。
- 专注于编程，而不是编程语言。我只注意包涵了Java的最小的可用子集，而忽略掉其他。

我需要一个标题，所以心血来潮选择了*How to Think Like a Computer Scientist*。

第一版教材很粗糙，但确实有效。学生们读完课本，懂得了足够的基础知识，以至我甚至可以利用课堂时间和他们一起讨论更难、更有趣的话题，并且（最重要的是）可以让学生们有足够的时间在课堂上做练习。

我将这本书按照GNU自由文档许可协议（GNU Free Documentation License）发布，让用户可以复制、修改和分发本书。

接下来发生了最酷的事情。Jeff Elkner，弗吉尼亚州的一位高中老师，使用了我的书，并且将其翻译成Python语言的版本。他寄给我他的翻译副本，于是我有了一次很奇特的经历——通过读我自己的书来学习Python。通过绿茶出版社（Green Tea Press），在2001年我出版了第一个Python版本。

2003年，我开始在欧林学院（Olin College）教学，并第一次需要教授Python语言。和Java的对比非常惊人。学生们困扰更少，学会得更多，从事更有意思的项目，总的来说得到了更多的乐趣。

结果就产生了本书，并使用了不那么宏伟堂皇的书名：*Think Python*。部分改动如下所述。

- 我在每章的结尾添加了一节关于调试的说明。这些章节描述寻找和避免bug的通用技巧，并警示Python中容易出错的误区。
- 我增加了更多的练习，小到简短的理解性测试，大到几个实际工程。并且编写了大部分练习的解答。
- 我添加了一系列案例研究——较长的示例，包括练习、解答以及讨论。其中有些是基于Swampy——我为Python课程所写的一套代码。Swampy、代码示例以及它们的解答，可以在<http://thinkpython.com>上找到。
- 我扩展了关于程序开发计划和基础设计模式的讨论。
- 我增加了关于调试、算法分析和使用Lumpy画UML图的附录。

我希望你喜欢这本书，并希望它至少能提供一点帮助，助你学会像计算机科学家那样编程和思考。

—— Allen B. Downey

Needham, MA

# 关于作者

Allen Downey是欧林工程学院的计算机科学教授。他曾在韦尔斯利学院、科尔比学院和加州大学伯克利分校教授计算机科学课程。他从加州大学伯克利分校获得计算机科学博士学位，并拥有MIT的硕士和学士学位。

# 关于封面

本书封面的动物是卡罗来纳鹦鹉，也叫卡罗来纳长尾鹦鹉（学名 *Conuropsis carolinensis*）。这种鹦鹉分布于美国东南部，最北到达纽约和大湖区，但主要分布在佛罗里达州到卡罗来纳州一带。

卡罗来纳鹦鹉主色是绿色，头部黄色，成熟时前额和两颊会出现一些橙红色的条纹。它的平均尺寸是31~33cm。它叫声狂暴而巨大，并且在捕食过程中会喋喋不休。

它居住在沼泽与河畔的树洞中。卡罗来纳鹦鹉是喜欢群居的生物，平时以小群体形式生活，在捕食时可以达到几百只。

不幸的是，这些捕食过程往往在农田的庄稼地里进行，农夫会射击它们，以免破坏庄稼。它们的群体特性让它们会集体救助受伤的鹦鹉，结果让农夫可以一次杀光整群鹦鹉。不但如此，它们的羽毛被用做妇女的帽饰，也有一些鹦鹉被作为宠物。这些因素组合起来，导致在19世纪晚期，卡罗来纳鹦鹉变得非常稀少，并且禽类疾病也加剧了它们的减少。到1920年左右，这个物种灭绝了。

今天，全世界的博物馆中保存了700多只卡罗来纳鹦鹉的标本。

封面图片来自《约翰逊的自然历史》（*Johnson's Natural History*）。

# 译后记

《像计算机科学家一样思考》这一系列书，早有耳闻，它可谓开创了程序设计入门书的一个新思路。授人以鱼，不若授人以渔；教人编程，不如引导人思考；教人语言细节，不若指明语言精要。而结合Python语言之后，得到的《像计算机科学家一样思考Python》这本书，则是在这个思路上走到了一个极致的佳作。

我是工作之后才开始接触Python的。那时候已经接触过C/C++、Java、C#等传统风格的语言，再看到Python，不免耳目一新。为何以前觉得难以理解的程序设计理念，在Python中的表达却这么简洁而易懂？为何以往需要好多行代码绞尽脑汁才能编写出来的功能，在Python里却只需要几个简单调用即可？为何繁复的集合操作，在Python中却只需要一行for循环语句就完成了？为何Python的文档那么容易找，还可以使用交互模式轻松尝试？每次使用Python编写程序之后，总会感慨，当初初学程序设计语言的时候，如果教的是Python该多好。相信所有学过C/C++之后再接触Python、Ruby、Haskell、Lisp等类似语言的人，都会有相同的感受吧。

那么是什么原因让C/C++几乎垄断了程序设计语言的教材呢？历史惯性。在计算机科学教育开始普及的20世纪70、80年代，C语言正在其鼎盛时期，几乎所有的人都在用C开发程序，操作系统、软件、游戏几乎都是用C甚至汇编开发的。硬件的限制，让那些更抽象、更高阶的语言，无法普及开来。因此教学自然也使用它。久而久之形成了惯性，到了新世纪，程序设计的教学已经赶不上语言发展的潮流了。我们的程序越来越复杂，越来越像人脑，而教学的语言仍然在使用最贴近机器语言的C。而C++、Java、C#不过是在这一惯性上多走了五十步而已。

本书正是扭转这种矛盾局面的一个有益的尝试。《像计算机科学家一样思考》是对程序设计教学模式的真谛的领悟，而使用Python这种简洁强大的高阶语言，也正是这种新思路最贴切的贯彻。授人以渔，自然应当用最好的渔具；引导人思考，当然也应使用更贴近人的思路而不是机器思路的语言。Python在高阶语言中，是一个从理念和实际综合考量后非常合适的候选。

在翻译过程中我发现，本书不但思路很贴切其教学主旨，从行文和用例来看也非常浅显易懂。全书讲了非常多的程序设计理念，在读过之后却会觉得那些理念都很自然，大概也是因为作者苦心安排，前后穿插，让读者能循序渐进地明白每个程序设计理念是因为什么出现的原因吧。这种风格，再配合上精心编辑的示例，用于介绍任何程序设计语言，都是非常合适的。

如果将来我的孩子愿意学习程序设计，我愿意用这本书教他。

尽管我已尽最大努力用心使译文准确、完善，但仍然难免有疏漏之处，如发现问题，欢迎批评指正。电子邮箱  
[zhaopuming@gmail.com](mailto:zhaopuming@gmail.com)。

# 关于译者

赵普明 清华大学计算机科学技术专业毕业，长期从事Web应用、高性能服务器以及计算平台的开发。从2.3版本开始接触Python，至今已逾5年。工作中使用Python编写脚本程序，用于快速原型构建以及日志计算等日常作业；业余时，作为一个编程语言爱好者，对D、Kotlin、Clojure、Scala、Ruby等语言均有了解，但至今仍为Python独特的风格、简洁的设计而惊叹。

# 致谢

非常感谢Jeff Elkner，他将我的Java书翻译成Python，致使我开始这个项目，并向我介绍了Python语言，结果成为我最爱的编程语言。

另外感谢Chris Meyers，他在*How to Think Like a Computer Scientist*一书中贡献了好几章。

感谢自由软件基金会（Free Software Foundation）开发了GNU自由文档协议，让我和Jeff以及Chris的合作成为可能。感谢创用CC（Creative Commons）开发了我们现在使用的协议。

感谢Lulu，负责*How to Think Like a Computer Scientist*的编辑。

感谢所有参与了本书早期版本编写的学生，以及所有（下面列出的）贡献者提供的修订和建议。

# 贡献者列表

在最近几年中，超过100名眼光犀利、思维敏捷的读者给我寄来了建议和修订。他们对这个项目的贡献和热情，对我是极大的帮助。如果你有建议或者修订意见，请发邮件到feedback@thinkpython.com。如果我根据你的回馈做出了修改，会将你加入到贡献者列表中（除非你要求被隐藏）。

如果你给出错误出现的位置的部分语句，会让我更容易搜索。页码或者章节号码也可以，但并不那么容易处理。谢谢！

- Lloyd Hugh Allen对8.4节提出了修订建议。
- Yvon Boulianne对第5章提出了一个语义错误的修订建议。
- Fred Bremmer对2.1节提出了一个修订建议。
- Jonah Cohen编写了Perl脚本将本书的LaTeX源码转换成美丽的HTML。
- Michael Conlon提出了第2章的一个语法错误，并提出第1章的格式改进，并且他开启了对解释器的技术讨论。
- Benoit Girard寄来一个对5.6节的有趣的修订。
- Courtney Gleason 和 Katherine Smith 编写了*horsebet.py*，在本书的早期版本中作为一个案例研究。他们的程序现在可以在网站上找到。
- Lee Harr提交了很多修订建议，我们没有空间在这里一一列出，并且他确实应当被列为本书的一位主要编辑。
- James Kaylin是一名使用本书的学生。他提交了许多修订。
- David Kershaw 修正了3.10节中错误的*catTwice*函数。
- Eddie Lam提出了第1、2、3章的很多修订建议，他也修正了*Makefile*，这样第一次运行时会自动建立索引。他也帮助我们设置了一个版本管理方案。
- Man-Yong Lee 寄来了对2.4节中的示例代码的修订。
- David Mayo指出第1章中的单词“unconsciously”需要被修改为“subconsciously”。
- Chris McAloon 寄来了对3.9节和3.10节的一些修订。
- Matthew J. Moelter是本书的长期贡献者，提出了很多修订建议。

- Simon Dicon Montford报告了第3章中缺失的函数定义以及几个错别字。他也发现了第13章中的increment函数的错误。
- John Ouzts修正了第3章中“返回值”的定义。
- Kevin Parks对关于本书如何分布提出了有价值的评论和建议。
- David Pool发来了第1章中术语表中的错别字，以及鼓励的赞美之言。
- Michael Schmitt寄来了关于文件和异常的章节的修订建议。
- Robin Shaw指出了13.1节中的一个错误，printTime函数在一个示例中没有定义就使用了。
- Paul Sleigh在第7章中找到一个错误，并发现了Jonah Cohen用于生成HTML的Perl脚本的bug。
- Craig T. Snydal在德鲁大学(Drew University)的一门课上试验这个课本，他提出了好几个有价值的建议和修订。
- Ian Thomas和他的学生们使用这本书作为编程课程的教材。他们第一个尝试使用本书后半部分的章节，并且提出了许多修正和建议。
- Keith Verheyden发来了第3章的一个修正。
- Peter Winstanley让我们知道了第3章的拉丁文中一个长期存在的错误。
- Chris Wrobel修正了文件I/O和异常一章的代码错误。
- Moshe Zadka对本书有不可估量的贡献。他编写了关于字典的一章的第一版草稿，并在本书的早期阶段持续提供指导。
- Christoph Zworschke发来了几个修正和教学法的建议，并解释了gleich和selbe的区别。
- James Mayer发送给我们非常多的拼写错误，包括贡献者列表中的两个错误。
- Hayden McAfee发现了两个示例之间潜在的冲突。
- Angel Arnal是翻译本书的西班牙语版本的国际团队的一员。他也发现了英文版中的几个错误。
- Tauhidul Hoque和Lex Berezhny创建了第1章中的图表，并改进了很多其他图表。
- Dr. Michele Alzetta发现了第8章中的一个错误，并发来了一些有趣的教学法评论，以及关于斐波那契数列和Old Maid的建议。
- Andy Mitchell发现了第1章中的一个录入错误，以及第2章中一个错误的示例。
- Kalin Harvey对第7章的一个说明提供了建议，并发现了几个录入错误。

- Christopher P. Smith发现了几个录入错误，并帮助我们更新本书到Python 2.2。
- David Hutchins发现了前言中的一个错别字。
- Gregor Lingl在奥地利维也纳的一个高中教授Python。他正在翻译本书的德文版，并发现了第5章中的几个错误。
- Julie Peters发现了前言中的一个错别字。
- Florin Oprina发来一个makeTime的改进，printTime的一个修正，以及发现的一个重要的录入错误。
- D. J. Webre对第3章的一个说明提出了建议。
- Ken在第8、9、11章中发现了好几个错误。
- Ivo Wever在第5章发现一个录入错误，并对第3章中的一个说明提出了建议。
- Curtis Yanko对第2章中的一个描述提出了建议。
- Ben Logan发来许多发现的录入错误，并发现了翻译HTML的问题。
- Jason Armstrong发现了第2章中一个漏掉的词。
- Louis Cordier发现了第16章中有一个代码和文本不一致的地方。
- Brian Cain在第2章和第3章中提出了几个描述的改进建议。
- Rob Black发来了许多修正，包括一些针对Python 2.2的修改。
- 巴黎中央理工大学的Jean-Philippe Rey发来了一些补丁，包括对Python 2.2的更新，以及其他一些细心的改进。
- 乔治华盛顿大学的Jason Mader提供了许多有用的建议和改正。
- Jan Gundtofte-Bruun提醒我们“a error”应改为“an error”。
- Abel David和Alexis Dinno提醒我们“matrix”的复数形式是“matrices”而不是“matrixes”。这个错误在书中已经存在了多年，但两个同姓的读者同一天报告了它。真的很奇怪。
- Charles Thayer鼓励我们删除掉一些语句结尾的分号，并建议我们理清“形参”和“实参”的使用。
- Roger Sperberg指出了第3章的一个逻辑错误。
- Sam Bull指出了第2章中一段令人困惑的描述。
- Andrew Cheung指出了两处“定义前先使用”的错误。
- C. Corey Capel发现了调试的第三定律中缺失的单词，以及第4章的一个录入错误。
- Alessandra帮助我们理清了一些关于Turtle的困惑。
- Wim Champagne在字典示例中发现一个错误。
- Douglas Wright在弧度计算中发现了一个除法向下取整的错误。
- Jared Spindor发现了一处句尾的无用词。

- Lin Peiheng发来了许多很有用的建议。
- Ray Hagtvedt发来了两处错误和一处不是那么错的错误。
- Torsten Hübsch指出Sawmpy中的一处不一致。
- Inga Petuhhov修正了第14章中的一个示例。
- Arne Babenhauserheide发来了几个有用的修正。
- Mark E. Casida非常善于发现重复的单词。
- Scott Tyler填上了一个缺失的“that”，并发来了一堆修正。
- Gordon Shephard发来了几个修正，每个都用单独的邮件。
- Andrew Turner发现了第8章中的一个错误。
- Adam Hobart修正了一个在弧度计算中除法向下取整的错误。
- Daryl Hammond和Sarah Zimmerman指出我过早提出了math.pi。并且Zim发现了一个录入错误。
- George Sass在调试章节中发现了一个bug。
- Brian Bingham建议了练习11–10。
- Leah Engelbert-Fenton指出我用tuple做变量名称，这正违反了我的建议。然后他发现了一堆录入错误以及一个“定义前先使用”。
- Joe Funke发现了一个录入错误。
- Chao-chao Chen在斐波那契示例中发现了一个不一致处。
- Jeff Paine知道space和spam的区别。
- Lubos Pintes发来一个录入错误。
- Gregg Lind和Abigail Heithoff建议了练习14–4。
- Max Hailperin发来了许多修正和建议。Max是非凡的《具体抽象》(Concrete Abstractions)一书的作者之一。在读完本书之后你可能会想要读那本书。
- Chotipat Pornavalai在一个错误信息中发现了一个错误。
- Stanislaw Antol寄来了一个很有用的建议列表。
- Eric Pashman对第4章到第11章发来了许多修正。
- Miguel Azevedo发现了一些录入错误。
- Jianhua Liu发来了一长列修正。
- Nick King发现了一个缺失单词。
- Martin Zuther发来了一长列建议。
- Adam Zimmerman发现了我举例的一个“实例”中的不一致处，以及其他一些错误。
- Ratnakar Tiwari建议加一个脚注说明什么是“退化”三角形。
- Anurag Goel提出了is\_abecedarian的另一个解答，并发来其他一些修正。他还知道如何拼写Jane Austen。

- Kelli Kratzer发现了一个录入错误。
- Mark Griffiths指出了第3章中的一个令人困惑的示例。
- Roydan Ongie发现了我的牛顿方法的一个错误。
- Patryk Wolowiec帮我解决了一个HTML版本的问题。
- Mark Chonofsky告诉我Python 3中的新关键字。
- Russell Coleman帮我修正了几何错误。
- Wei Huang发现了几处录入错误。
- Karen Barber发现了本书中最古老的录入错误。
- Nam Nguyen发现了一个录入错误，并指出我使用了装饰器模式但并没有用它的名字。
- Stéphane Morin发来了一些建议和修正。
- Paul Stoop修改了一个`uses_only`中的录入错误。
- Eric Bronner指出了关于操作符顺序的讨论中的一个困惑之处。
- Alexandros Gezerlis提交的建议的数量和质量都设置了一个新的标准。我们非常感谢他！
- Gray Thomas知道哪边是左哪边是右。
- Giovanni Escobar Sosa发来一长列的修正和建议。
- Alix Etienne修正了一个URL。
- Kuang He发现一个录入错误。
- Daniel Neilson修正了一个关于操作符顺序的错误。
- Will McGinnis指出`polyline`在两个地方定义的不同。
- Swarup Sahoo发现了一个缺失的分号。
- Frank Hecker指出一个练习不细致，并发现了几个坏链接。
- Animesh B帮助我清理了一个令人困惑的示例。
- Martin Caspersen发现了两处取整错误。
- Gregor Ulm发来一些修正和建议。

# 0' Reilly Media, Inc. 介绍

0' Reilly Media通过图书、杂志、在线服务、调查研究和会议等方式传播创新知识。自1978年开始，0' Reilly一直都是前沿发展的见证者和推动者。超级极客们正在开创着未来，而我们关注真正重要的技术趋势——通过放大那些“细微的信号”来刺激社会对新科技的应用。作为技术社区中活跃的参与者，0' Reilly的发展充满了对创新的倡导、创造和发扬光大。

0' Reilly为软件开发人员带来革命性的“动物书”；创建第一个商业网站（GNN）；组织了影响深远的开放源代码峰会，以至于开源软件运动以此命名；创立了Make杂志，从而成为DIY革命的主要先锋；公司一如既往地通过多种形式缔结信息与人的纽带。0' Reilly的会议和峰会集聚了众多超级极客和高瞻远瞩的商业领袖，共同描绘出开创新产业的革命性思想。作为技术人士获取信息的选择，0' Reilly现在还将先锋专家的知识传递给普通的计算机用户。无论是通过书籍出版，在线服务或者面授课程，每一项0' Reilly的产品都反映了公司不可动摇的理念——信息是激发创新的力量。

## 业界评论

“0' Reilly Radar博客有口皆碑。”

——Wired

“0' Reilly凭借一系列（真希望当初我也想到了）非凡想法建立了数百万美元的业务。”

——Business 2.0

“0' Reilly Conference是聚集关键思想领袖的绝对典范。”

——CRN

“一本0' Reilly的书就代表一个有用、有前途、需要学习的主题。”

——Irish Times

“Tim是位特立独行的商人，他不光放眼于最长远、最广阔的视野并且切实地按照Yogi Berra的建议去做了：‘如果你在路上遇到岔路口，走小路（岔路）。’回顾过去Tim似乎每一次都选择了小路，而且有几次都是一闪即逝的机会，尽管大路也不错。”

——Linux Journal

# 第1章 程序之道

本书的目标是教会你像计算机科学家一样思考。这种思考方式综合了数学、工程学以及自然科学的一些最优秀的特性。计算机科学家和数学家类似，他们使用形式语言来描述理念（特别是计算）；和工程师类似，他们设计产品，将元件组装成系统，评估选择不同的方案；和自然科学家类似，他们观察复杂系统的行为，形成科学假设，并检验其预测。

计算机科学家最重要的技能是问题解决。问题解决意味着发现问题，创造性地思考解决方案，以及清晰准确地表达解决方案的能力。事实证明，学习编程的过程，是训练问题解决能力的绝佳机会。这也是为什么本章标题是“程序之道”的原因。

一方面，你将学会编程，其本身就是一个非常有用的技能；另一方面，你可以使用编程作为工具，去达到更高的目标。随着本书的深入，那个目标会逐渐明晰。

## 1.1 Python编程语言

你将学习的编程语言是Python。Python是一种高级语言。你可能还听说过其他的高级编程语言，比如C、C++、Perl和Java。

另外，也有低级语言，有时被称为“机器语言”或者“汇编语言”。一般地讲，计算机只能运行低级语言编写的程序。所以，使用高级语言编写的程序必须先处理过才能运行。额外的处理过程需要花费一点时间，这也是高级语言的一个小缺点。

但高级语言的优点是巨大的。首先，使用高级语言编写程序容易得多。高级语言的程序编写时耗时更少，程序更短，更容易阅读，且更容易保证正确。其次，高级语言是可移植的，这意味着仅需稍微修改或者不用修改，它们就可以在不同类型的计算机上运行。而低级语言的程序只能在一种计算机上运行，想用在不同的机器上，必须重写。

因为这些优点，几乎所有的程序都是用高级语言编写的。低级语言只在很稀少的特殊场景中使用。

有两种程序可以处理高级语言并将其转换为低级语言：解释器和编译器。解释器读入一段高级语言程序，并执行它。也就是说，它会按照程序的指令运行。它每次处理一小部分程序，交替读入代码行并进行运算。图1-1显示了解释器的结构。

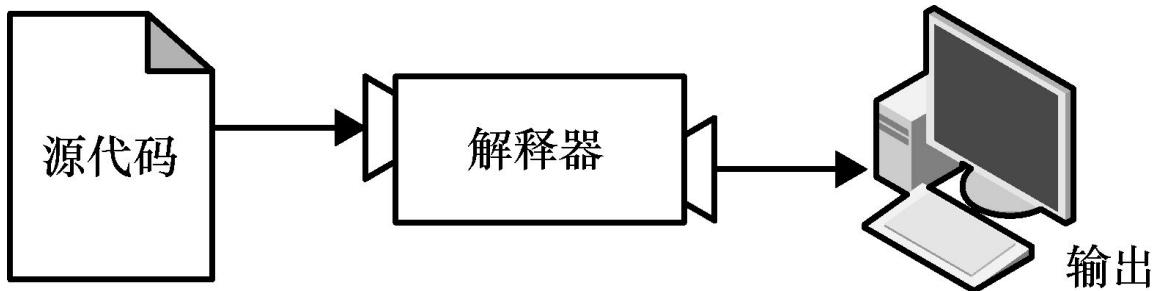


图1-1 解释器每次处理一小部分程序，交替读入代码行并进行运算

编译器读入程序，将其完整地编译为低级语言，才能运行。在这个场景中，高级语言的程序称为源代码，编译而成的程序称为目标代码或可执行代码。一旦程序编译完成，就不需要再进行编译，直接可以重复执行。图1-2显示了一个编译器的结构。

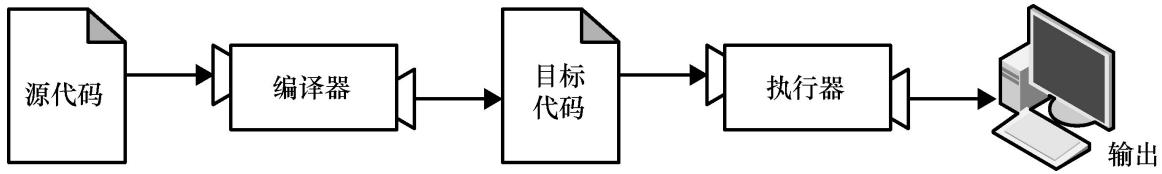


图1-2 编译器将源代码编译为可以在硬件上执行的目标代码

Python被认为是解释语言，因为它是使用解释器执行的。解释执行的方式有两种：交互模式和脚本模式。在交互模式下，用户输入一段Python代码，解释器显示结果：

```
>>> 1 + 1  
2
```

这里>>>形的符号，用来提示用户解释器已经准备好接收代码输入。如果你输入 $1+1$ ，解释器会答复2。

或者，你可以把代码保存到一个文件里，并使用解释器来执行文件内容。这样的文件称为脚本。依照惯例，Python脚本文件的后缀是.py。

要想执行脚本，需要将文件名通知给解释器。如果你有一个名为dinsdale.py的脚本，并且使用UNIX命令行，则可以输入python dinsdale.py。在其他开发环境中，脚本执行的细节会有所不同。你可以在Python网站<http://python.org>上查看自己使用的环境的具体说明。

使用交互模式对测试小段代码很方便，因为你可以输入代码并立即执行。但代码超过一定行数时，就应当保存到脚本文件中，以便于未来进行修改和执行。

## 1.2 什么是程序

程序是指一组定义如何进行计算的指令的集合。这种计算可能是数学计算，例如解方程组或者查找多项式的根，但也可以是符号运算，例如搜索和替换文档中的文本，或者（很奇怪地）编译一个程序。

在不同的语言中，程序的细节有所不同，但几乎所有语言都会出现几类基本指令。

- 输入：从键盘、文件或者其他设备中获取数据。
- 输出：将数据显示到屏幕或者发送到文件或其他设备中。
- 数学：进行基本数学操作，比如加法或乘法。
- 条件执行：检查某种条件的状态，并执行相应的代码。
- 重复：重复执行某种动作，往往在重复中有一些变化。

信不信由你，这差不多就是全部了。你所遇到过的所有程序，不论多么复杂，也都是由类似上面的这些指令组成的。所以我们可以把编程看做将大而复杂的任务分解为更小的子任务的过程，依此分解，直到任务简单得可以由上面的这些指令组合完成。

这看起来也许有些模糊，但我们在讨论算法时再回到这个话题上来。

## 1.3 什么是调试

程序是很容易出错的。因为某种古怪的原因，程序错误被称为bug，而查捕bug的过程称为调试（debugging）。

一个程序中可能出现3种类型的错误：语法错误、运行时错误和语义错误。对它们加以区分，可以更快地找到错误。

## 1.4 语法错误

Python程序在语法正确的情况下才能运行；否则，解释器会显示一条错误信息。语法指的是程序的结构以及此结构的规则。比如，括号必须前后匹配，所以`(1+2)`是合法的，而`8)`就是一个语法错误。

在英语中，读者可以容忍大多数语法错误，因此我们可以阅读E. E. Cummings的诗，而不需要发出错误信息。但Python并不如此宽容。程序中只要出现一处语法错误，Python就会显示错误信息并退出，结果你的程序就无法运行了。在编程生涯的最初几周中，可能会需要花费大量时间来查找语法错误。但随着经验的增加，犯错减少，查找起来也会更快。

## 1.5 运行时错误

第二类错误是运行时错误，这样称呼是因为这种错误只有程序运行后才会出现。这些错误也常被称为异常，因为它们常常表示某些异常的（而且不好的）事情发生了。

运行时错误在开头几章中的简单示例里很少会出现，所以可能要过一段时间，你才会遇到一个。

## 1.6 语义错误

第三类错误是语义错误。如果你的程序中有一个语义错误，程序仍会成功运行，而不会产生任何错误信息，但是它不会执行正确的逻辑。它会做其他的事情。特别地，它会做的正是你告诉它所做的。

这里的问题在于你写出的代码和你想要写的代码并不一致。程序的意思（语义）是错的。查找语义错误会很麻烦，因为需要你反向查找，查看程序输出并尝试弄明白它到底做了什么。

## 1.7 实验型调试

你将会掌握的一个最重要的技能就是调试。虽然调试可能较困惑，但它的确是编程活动中最动头脑、最有挑战、最有趣的部分。

在某种程度上，调试和刑侦工作很像。你会面对一些线索，而且必须推导出事情发生的过程，以及导致现场结果的事件。

调试也像是一种实验科学。一旦你猜出错误的可能原因，就可以修改程序，再运行一次。如果你猜对了，那么程序的运行结果会符合你的预测，这样就离正确的程序更近了一步。如果你猜错了，则需要重新思考。正如夏洛克·福尔摩斯所说的：“当你排除掉所有的可能性，那么剩下的，不管多么不可能，必定是真相。”（柯南·道尔《四签名》）

对某些人来说，编程和调试是同一件事。也就是说，编程正是不断调试修改直到程序达到设计目的的过程。这种想法的要旨是，你应该从一个能做某些事的程序开始，然后做一点点修改，并调试修改，如此迭代，以确保总是有一个可以运行的程序。

例如，Linux是包含了成千上万行代码的操作系统，但最开始只是Linus Torvalds 编写的用来研究Intel 80386芯片的简单程序。据Larry Greenfield所说，“Linus最早的一个程序是交替打印AAAA和BBBB。后来这些程序演化成了Linux。”（《Linux用户指南Beta版本1》）

后面的章节会介绍更多关于调试的技巧，以及其他编程实践技巧。

# 1.8 形式语言和自然语言

自然语言是指人们所说的语言，比如英语、西班牙语和法语。它们不是由人设计而来的（虽然人们会尝试加以语法限制），而是自然演化而来的。

形式语言则是人们设计出来用于特别用途的语言。比如，用来表达数学公式的符号标注系统，是一种特别擅于表示数字和符号的关系的形式语言。化学家则使用一种形式语言表达分子的化学结构。并且最重要的是：

编程语言是人们设计出来用来表达计算过程的形式语言。

形式语言倾向于对语法做出严格的限制。比如， $3+3=6$ 是语法正确的数学表达式，但 $3+ = 3\$6$ 则不是。 $H_2O$ 是语法正确的化学方程，而 $_2Zz$ 则不是。

语法规则有两种，分别适用于记号（token）和结构。记号是语言的基本元素，比如词、数字和化学元素。 $3+ = 3\$6$ 的一个问题就是\$在数学表达式中（至少就我所知）不是合法记号。相似地， $_2Zz$ 不合法，是因为并不存在缩写为Zz的化学元素。

第二种语法错误是关于语句的结构的。也就是说，记号所排列的方式。语句 $3+ = 3$ 不合法，因为虽然+和=是合法记号，但不能将它们连续放置。相似地，在化学表达式里，下标数字应该出现在元素名称之后，而不是之前。

## 练习1-1

写一个结构良好的英语句子，但其中包含非法的记号。再写一个句子，包含的记号都是合法的，但结构不合法。

当你阅读英语的句子或形式语言的语句时，你需要弄清句子的结构是什么（虽然在自然语言中这个过程是下意识完成的）。这个过程称为语法分析。

比如，当你听到这句话：“硬币掉了。”会理解到，“硬币”是主语，而“掉了”是谓语。一旦解析完一句话，就能弄清楚它的涵义，或者说语义。假设你知道硬币是什么，并知道掉了意味着什么，就能够理解这句话的涵义。

虽然形式语言和自然语言有很多共同的特点——记号、结构、语法以及语义——它们也有一些区别。

- **歧义性**: 自然语言充满了歧义，人们通过上下文线索和其他信息来处理。形式语言通常设计为几乎或者完全没有歧义，意即不论上下文环境如何，任何表达式都只有一个意义。
- **冗余性**: 为了弥补歧义，减少误解，自然语言采用大量的冗余。因此，它们常常冗长繁复。形式语言则不那么冗余，相对紧凑。
- **字面直接性**: 自然语言充满了成语和比喻。比如有人说，“硬币掉了”，并不一定是硬币，也不一定是什么掉了（“The penny dropped”在英语里的意思是：经过一段困惑后，突然发现之前没有意识到的事情）。形式语言则严格按照它字面的意思表达涵义。

说着自然语言长大的人（所有人）都需要经历一段挣扎才能适应形式语言。在某种意义上，形式语言和自然语言的区别，与诗词和散文的区别类似，而且更甚。

- **诗词**: 字词的使用，既考虑到它们的音韵，也考虑到它们的意义，而整首诗合起来表达某种意境或情绪反应。歧义不仅常见，而且常常是刻意为之。
- **散文**: 字词的意义更加重要，而且句子的结构也提供更多的意义。散文比诗词更容易分析，但仍然有不少歧义。
- **程序**: 计算机程序的意义不含歧义，而且直接如字面所指。并且可以完全通过它的记号和结构理解其意义。

阅读代码（以及其他形式语言）时有如下建议。首先，记着形式语言的密度远远大于自然语言，所以阅读它们需要花费更多时间。其次，文法结构非常重要，所以直接自顶而下、从左至右的阅读顺序，并不一定是最好的。相反，试着学会在头脑中解析程序，辨别出记号并解析出结构。最后，细节很重要。在自然语言中常常可以忽略的小错误，如拼写或者标点错误，在形式语言中往往影响重大。

## 1.9 第一个程序

依照传统，学习一门新语言，写的第一个程序都叫“Hello, World!”，因为这个程序所做的事情就是显示“Hello, World!”。在Python中，它是这个样子：

```
print 'Hello, World!'
```

这是print语句的一个示例。print并不会真地往纸上打印文字，而是在屏幕上输出值。这个例子输出的结果是：

```
Hello, World!
```

程序中的引号表示输出的文本的开始和结束；在输出结果中它们并不显示。

在Python 3中，这个语句的语法略有不同：

```
print('Hello, World!')
```

括号表示print是一个函数。我们将在第3章中讨论函数的话题。

在本书后面的内容中，我会使用print语句。如果你是使用Python 3，需要自己转换。但除了这一处之外，很少有其他需要担心的区别了。

## 1.10 调试

在电脑前阅读本书会是一个好主意，因为你可以边看边试验书中的示例。大部分的代码都可以使用交互模式运行。当然，如果你把代码存入到脚本中，可以方便尝试一些小修改。

每当你试验新的语言特性时，应当试着故意犯错。比如，在“Hello, World!”程序中，如果少写一个引号，会发生什么？如果两个引号都不写，会怎么样？如果把print拼写错了，会如何？

这种试验会帮助你记住所读的内容，也能帮你学会调试，因为这样能看到不同的错误信息代表着什么。现在故意犯错，总比今后在编码中意外出错好。

编程，特别是调试，有时候会引发强烈的情绪。如果你挣扎于一个困难的问题，可能会感觉到愤怒、沮丧以及窘迫。

有证据表明，人们会像对待人一样对待电脑。当电脑良好完成工作时，我们会把它们当作队友，而当它们难以控制、粗暴无礼的时候，我们会按照对待那些粗暴固执的人一样对待它们（《媒体等同：人们该如何像对待真人实景一样对待电脑、电视和新媒体》，Reeves和Nass著）。

对这些反应行为有所准备，可能会帮助你更好地对待电脑。一种方法是把它当作你的雇员，它有一定的长处，比如速度和精度，也有特定的弱点，比如没有同情心，以及无法顾全大局。

你的任务是做一个好经理：设法扬长避短。并找到方法控制你的情绪去面对问题，而不是让你的反应影响工作的效率。

学习调试可能会带来挫折感，但它是一个有价值的技能，并在编程之外还有很多用途。每章的结尾处都有一节类似于本节的关于调试技巧的讨论。希望它们能带来帮助！

## 1.11 术语表

问题解决 (problem solving)：总结问题、寻找解决方案以及表达解决方案的过程。

高级语言 (high-level language)：设计来方便人们读写的编程语言，比如Python。

低级语言 (low-level language)：设计来方便计算机执行的编程语言，也被称为“机器语言”或“汇编语言”。

可移植性 (portability)：程序的一种属性：可以在多种类型的计算机上运行。

解释 (interpret)：按照一行一行解释翻译的方式来执行高级语言编写的程序。

编译 (compile)：一次性将一个高级语言编写的程序翻译为低级语言程序，之后可以单独运行。

源代码 (source code)：使用高级语言编写的程序，在编译之前称为源代码。

目标代码 (object code)：编译器输出的程序。

可执行文件 (executable)：目标代码的另一个名字，表示它可以直接受执行。

提示符 (prompt)：解释器显示出来的文字，用来表示它准备好接收用户的新的输入。

脚本 (script)：保存在文件中的程序（用于被解释器解释执行）。

交互模式 (interactive mode)：使用Python解释器的一种方式。在解释器的提示处输入命令和表达式。

脚本模式 (script mode)：使用Python解释器的另一种方式。读取并执行一个脚本文件中的代码。

程序 (program)：一系列代码指令的集合，指定一种运算。

算法 (algorithm) : 解决某一类问题的通用运算流程。

bug: 程序中的一个错误。

调试 (debugging) : 发现和解决程序中出现的3类错误的过程。

语法 (syntax) : 程序的结构。

语法错误 (syntax error) : 程序中的一种错误，导致它无法进行语法解析（因此也无法被解释器执行）。

异常 (exception) : 程序运行中发现的错误。

语义 (semantics) : 程序表达的意义。

语义错误 (semantic error) : 程序中的一种错误，导致它运行所做的事情和程序员设想的不同。

自然语言 (natural language) : 自然演化而来的人们所说的语言。

形式语言 (formal language) : 人们设计来用于某些特定目的的语言，例如表达数学概念或者计算机程序。所有的编程语言都属于形式语言。

记号 (token) : 程序的语法结构的最基本单位，类似于自然语言中的词。

语法分析 (parse) : 检查程序并分析其语法结构。

print语句 (print statement) : 一个指令，可以通知Python解释器在屏幕上输出一个值。

## 1.12 练习

### 练习1-2

使用浏览器访问Python的网站<http://python.org>。这个页面包含了Python的相关信息，以及到Python相关页面的链接，并且可以在上面搜索Python的文档。

### 练习1-3

启动Python解释器，输入`help()`来启动在线帮助功能。或者，可以输入`help('print')`获取`print`语句的相关文档。

如果这个例子不能运行，可能是由于需要安装额外的Python文档模块，或者需要设置环境变量；详细情况依赖于你使用的操作系统和Python版本。

### 练习1-4

启动Python解释器，把它当作计算器使用。Python的数学运算语法和标准的数学标记法一致。比如，和预料一样，符号`+`、`-`和`/`分别表示加法、减法和除法。乘法的符号是`*`。

如果你用43分30秒跑完10公里，那么你平均跑1英里需要多长时间？平均速度是多少英里每小时？（提示：1英里相当于1.61公里。）

# 第2章 变量、表达式和语句

## 2.1 值和类型

值 (value) 是程序操作的最基本的东西。比如一个字符或者数字。至今我们见过这些值: 1, 2, 以及'Hello, World!'。

这些值属于不同的类型: 2是一个整数, 而'Hello, World!'是一个字符串, 这么称呼是因为它包含了一“串”字符。你(和解释器)可以通过包含它们的引号确定字符串。

如果不确认一个值的类型, 解释器可以告诉你:

```
>>> type('Hello, World!')
<type 'str'>
>>> type(17)
<type 'int'>
```

不足为奇, 字符串的类型是'str', 而整数属于类型'int'。但并不明显的是, 包含一个小数点的数, 属于称为浮点型 ('float') 的类型, 因为这些数字的表达形式中有一个浮点 (floating-point)。

```
>>> type(3.2)
<type 'float'>
```

那么'17'和'3.2'这样的值呢? 它们看起来像是数字, 但又使用字符串常用的引号包含。

```
>>> type('17')
<type 'str'>
>>> type('3.2')
<type 'str'>
```

它们是字符串。

当输入一个很大的数字时, 你可能会忍不住想在数字中间加上逗号, 就像1,000,000这样。在Python中这并不是合法的数字, 但它是合法的表达式:

```
>>> 1,000,000  
(1, 0, 0)
```

当然，这和我们设想的完全不同！Python把`1,000,000`解释成一个用逗号分隔的数字序列。这是我们遇到的第一个语义错误：代码并不报错，可以正常运行，但是它做的并不是“正确”的事情。

## 2.2 变量

编程语言最强大的功能之一是操纵变量的能力。变量是指向一个值的名称。

赋值语句可以建立新的变量，并给它们赋值：

```
>>> message = 'And now for something completely different'  
>>> n = 17  
>>> pi = 3.1415926535897932
```

这个例子有3个赋值。第一个将一个字符串赋给叫做message的变量；第二个将17赋值给n；第三个将π的（近似）值赋给变量pi。

在纸上表达变量的一个常见方式是写下名称，并用箭头指向其值。这种图称为状态图，因为它显示了每个变量所在的状态（请将它看作变量的心理状态）。图2-1显示了前面例子的状态图。

```
message → 'And now for something completely different'  
n → 17  
pi → 3.1415926535897932
```

图2-1 状态图

变量的类型和它所指向的值类型相同。

```
>>> type(message)  
<type 'str'>  
>>> type(n)  
<type 'int'>  
>>> type(pi)  
<type 'float'>
```

### 练习2-1

如果输入一个以0开头的整数，则会得到一个令人困惑的错误：

```
>>> zipcode = 02492  
^
```

```
SyntaxError: invalid token
```

其他的数字又似乎可以正常运行，但结果也很奇怪：

```
>>> zipcode = 02132  
>>> zipcode  
1114
```

你能想出是怎么回事吗？提示：显示01、010、0100和01000的值。

## 2.3 变量名称和关键字

程序员常常选择有意义的名称作为变量名——以此标记变量的用途。

变量名可以任意长短。它可以包含字母和数字，但必须以一个字母开头。使用大写字母是合法的，但变量名使用小写字母开头是个好主意（后面你会看到为何如此）。

下划线“\_”可以出现在变量名称中。它经常出现在由多个词组成的变量名中，如my\_name或airspeed\_of\_unladen\_swallow。

如果你给变量取非法的名称，会得到一个语法错误：

```
>>> 76trombones = 'big parade'  
SyntaxError: invalid syntax  
>>> more@ = 1000000  
SyntaxError: invalid syntax  
>>> class = 'Advanced Theoretical Zymurgy'  
SyntaxError: invalid syntax
```

76trombones非法，因为它不是以字母开头。more@非法，是因为它包含了一个非法字符@。但class有什么问题？

原因是class是Python的一个关键字。解释器通过关键字来识别程序的结构，并且它们不能用来作为变量名称。

Python 2共有31个关键字：

and	del	from	not	while
as	elif	global	or	with
assert	else	if	pass	yield
break	except	import	print	
class	exec	in	raise	
continue	finally	is	return	
def	for	lambda	try	

在Python 3中，exec不再是关键字，但nonlocal是一个新的关键字。

你可能需要随身携带这个列表。如果编译器抱怨你的变量名称而不知为何，先查看一下是不是在这个列表中。

## 2.4 操作符和操作对象

操作符是一些特殊符号，表示像加法乘法这样的运算。操作符所操作的值，称为操作对象。

操作符+、-、\*、/以及\*\*分别进行加法、减法、乘法、除法以及乘方操作。下面是示例：

```
20+32    hour-1    hour*60+minute    minute/60    5**2    (5+9)*(15-7)
```

在其他语言中，乘方操作用<sup>^</sup>表示，但是在Python中<sup>^</sup>用来表示按位异或（XOR）的操作。本书中不会涉及按位操作，但你可以在<http://wiki.python.org/moin/BitwiseOperators>中读到相关信息。

在Python 2中，除法可能和你想象的不一样：

```
>>> minute = 59
>>> minute/60
0
```

minute的值是59，而在传统的算术中，59除以60是0.98333，而不是0。这其中的差别是因为Python使用的是舍去式除法（floor division）。当两个操作对象都是整数时，结果也是个整数。舍去式除法会舍去小数部分，所以在这个例子中，结果被取整成为0。

在Python 3中，这个除法的结果是浮点数类型。而使用一个新的//操作符来表示舍去式除法。

如果两个操作数中有一个是浮点数，那么Python就会进行浮点除法，结果也是浮点数：

```
>>> minute/60.0
0.9833333333333328
```

## 2.5 表达式和语句

表达式是值、变量和操作符的组合。单独一个值也被看作一个表达式，单独的变量也是如此。所以下面都是合法的表达式（假设变量x已经被赋值）：

```
17  
x  
x + 17
```

语句是Python解释器能运行的一个代码单元。我们已经看到了两种语句：print语句和赋值语句。

从技术上说表达式也是一个语句，但把它们看作不同的事物区分开也许更简单。重要的区别是，表达式有值，而语句没有。

## 2.6 交互模式和脚本模式

使用解释型语言的好处之一在于你可以在交互模式中尝试一些代码片段，再将它们写入到脚本中。但交互模式和脚本模式也有一些不同，有时候会带来困惑。

比如，如果使用Python作为计算器，你可能会输入：

```
>>> miles = 26.2
>>> miles * 1.61
42.182
```

第一行给变量miles赋值，但没有可见的效果。第二行是一个表达式，所以解释器对其进行求值，并显示结果。于是我们知道马拉松的长度大概是42公里。

但如果将上面同样的代码写入到脚本中并运行，则得不到任何输出。在脚本模式中，一个单独的表达式，也是没有可见效果的。Python实际上会对表达式进行求值，但不会显示其结果。除非你叫它这么做：

```
miles = 26.2
print miles * 1.61
```

这种现象一开始可能会让人迷惑。

脚本通常包含一系列的语句。如果语句超过一行，那么会随着语句执行的顺序一行行显示结果。

比如，脚本

```
print 1
x = 2
print x
```

产生如下结果

```
1
2
```

赋值语句不会产生任何输出。

## 练习2-2

在Python解释器中输入下面的语句，看它们做了什么：

```
5  
x = 5  
x + 1
```

现在把同样的语句存入到一个脚本文件并运行。输出是什么？修改脚本，将所有的表达式都转换成print语句，再运行一遍。

## 2.7 操作顺序

当一个表达式中出现多个操作符时，求值的顺序依赖于优先级规则。对数学操作符，Python遵守数学的传统规则。缩略词PEMDAS可以帮助记忆这些规则：

- 括号（P, Parentheses）拥有最高的优先级，并可以用来强制表达式按照你需要的顺序进行求值。因为括号中的表达式会先执行，所以 $2 * (3 - 1)$ 的结果是4，而 $(1 + 1) ** (5 - 2)$ 的结果是8。你也可以利用括号使得表达式更加易读，就像 $(\text{minute} * 100) / 60$ 这样，即使这里增加括号并不会改变结果。
- 乘方（E, Exponentiation）操作拥有次高的优先级，所以 $2 ** 1 + 1$ 的结果是3，而不是4，而且 $3 * 1 ** 3$ 的结果是3，而不是27。
- 乘法（M, Multiplication）和除法（D, Division）优先级相同，并且高于亦有相同优先级的加法（A, Addition）和减法（S, Subtraction）。所以 $2 * 3 - 1$ 是5，而不是4，并且 $6 + 4 / 2$ 是8，而不是5。
- 优先级相同的操作按照自左向右的顺序求值（除了乘方以外）。所以表达式 $\text{degrees} / 2 * \text{pi}$ ，除法在乘法之前执行，结果乘以 $\text{pi}$ 。如果想除以 $2\pi$ ，可以使用括号，或者写为 $\text{degrees} / 2 / \text{pi}$ 。

其他操作符的优先级，我并不会花太多功夫记下来。如果只看表达式不能确定的话，使用括号指明优先级即可。

## 2.8 字符串操作

通常来说，字符串不能进行数学操作。即使看起来像数字也不行。下面的操作是非法的：

```
'2' - '1'      'eggs'/'easy'    'third'*'a charm'
```

字符串可以使用操作符`+`，但其功能可能和你想的不一样：它会进行拼接（concatenation）操作，意即将前后两个字符串首尾连接起来。比如：

```
first = 'throat'  
second = 'warbler'  
print first + second
```

这段程序的输出是`throatwarbler`。

操作符`*`也适用于字符串；它进行重复操作。比如，`'Spam'*3`的结果是`'SpamSpamSpam'`。如果`*`的两个操作对象之一是字符串，那另一个必须是整数。

字符串的`+`和`*`的应用，实际上和数字的加法与乘法类似。就像`4*3`与`4+4+4`相等一样，我们预期`'Spam'*3`与`'Spam'+'Spam'+'Spam'`也相等，实际也确实如此。另一方面，字符串的拼接与重复操作和整数的加法与乘法操作也有很大的不同。你能够想出加法的一个属性，字符串拼接操作并不支持吗？

## 2.9 注释

当程序变得更大更复杂时，读起来也更困难。形式语言很紧凑，经常会遇到一段代码，却很难弄清它在做什么、为什么那么做。

因此，在程序中加入自然语言的笔记来解释程序在做什么，是个好主意。这种笔记被称为注释（comments），它们以#开头：

```
# compute the percentage of the hour that has elapsed  
percentage = (minute * 100) / 60
```

在这个例子里，注释单独占据一行。你也可以把注释放到代码行的结尾：

```
percentage = (minute * 100) / 60      # percentage of an hour
```

从#开始到行尾的注释内容都会被解释器忽略掉——它们对程序本身运行没有任何影响。

注释最重要的用途在于解释代码并不显而易见的特性。我们可以合理地认为读者可以看懂代码在做什么，因此使用注释来解释为什么这么做，要有用得多。

下面这段注释与代码重复，毫无用处：

```
v = 5          # 将5赋值给v
```

而下面这段注释则包含了代码中看不到的有用信息：

```
v = 5          # 速度，单位是米/秒
```

选择好的变量名称，可以减少注释的需要，但长名字也会让复杂表达式更难阅读，所以这两者之间需要衡量舍取。

## 2.10 调试

至此你最可能犯的语法错误应该是使用非法变量名，比如错用关键字class和yield，或者包含了非法字符的odd~job和US\$。

如果在变量名中放一个空格，Python会将它当做缺少了操作符的两个操作对象：

```
>>> bad name = 5  
SyntaxError: invalid syntax
```

对语法错误来说，错误信息并没什么帮助。最常见的错误信息就是SyntaxError: invalid syntax和SyntaxError: invalid token，两者都没包含多少有用信息。

最可能犯的运行时错误应该是“没定义就使用”（use before def）。也就是，在给一个变量赋值之前，就直接使用它了。这种错误也可能在你拼写错误变量名的时候发生：

```
>>> principal = 327.68  
>>> interest = principle * rate  
NameError: name 'principle' is not defined
```

变量名是大小写敏感的，所以LaTeX和late $\backslash$ x并不相同。

这时候语义错误最可能的因素是弄错操作的顺序。比如，要计算 $1/2\pi$ ，你可能会这么写：

```
>>> 1.0 / 2.0 * pi
```

但除法会先进行，所以你得到的结果是 $\pi/2$ ，和预想不同！Python不可能得知你想要写什么，所以这里不会出现任何错误信息，只是会得到错误的计算结果。

## 2.11 术语表

值 (value)：程序中操作的数据的基本单位，例如数字或字符串。

类型 (type)：值的类别。我们现在见到的类型有整数 (int)，浮点数 (float)，以及字符串 (str)。

整数 (integer)：用来表达整数的类型。

浮点数 (floating-point)：用来表示带小数部分的数的类型。

字符串 (string)：用来表示一串字符的类型。

变量 (variable)：引用一个值的名字。

语句 (statement)：表示一个命令或行动的一段代码。至今我们见过赋值语句和print语句。

赋值 (assignment)：将一个值赋值给变量的语句。

状态图 (state diagram)：用来展示一些变量以及其值的图示。

关键字 (keyword)：编译器或解释器保留的词，用于解析程序；变量名不能使用关键字，如if, def, while等。

操作符 (operator)：用来表示简单的运算的特殊符号，如加法、乘法或者字符串串联。

操作数 (operand)：操作符所操作的值。

舍去式除法 (floor division)：两个数相除，并舍取结果的小数部分的操作。

表达式 (expression)：变量、操作符和值的组合，可以表示一个单独的结果值。

求值 (evaluate)：对表达式按照操作的顺序进行计算，求得其结果值。

优先级规则 (rules of precedence) : 当有多个操作符和操作对象时，用于指导操作计算顺序的规则。

拼接 (concatenate) : 将两个操作数首尾相连。

注释 (comment) : 代码中附加的注解信息，用于帮助其他程序员阅读代码，并不影响程序的运行。

## 2.12 练习

### 练习2-3

假设我们运行下面的赋值语句：

```
width = 17
height = 12.0
delimiter = '..'
```

对下面的表达式，给出其值和值的类型（即表达式的类型）：

1. width/2
2. width/2.0
3. height/3
4. 1 + 2 \* 5
5. delimiter \* 5

使用Python解释器来验证你的答案。

### 练习2-4

把Python解释器当做计算器来进行练习。

1. 半径为 $r$ 的球体的体积是 $(4/3) \pi r^3$ 。半径为5的球体体积是多少?  
提示：392.7是错的！
2. 假设一本书的定价是24.95美元，但是书店打了40%的折扣（6折）。运费是一本3美元，每加一本加75美分。60本书的总价是多少？
3. 如果我6:52时离开家，并以慢速（8分15秒/英里）跑1英里，接下来以7分12秒的速度跑3英里，再以慢速跑1英里。请问我回家吃早餐是什么时候？

# 第3章 函数

## 3.1 函数调用

在程序设计中，函数是指用于进行某种计算的一系列语句的有名称的组合。定义一个函数时，需要指定函数的名称并写下一系列程序语句。之后，就可以使用名称来“调用”这个函数。前面我们已经见过函数调用的例子：

```
>>> type(32)
<type 'int'>
```

这个函数的名称是type，括号中的表达式我们称之为函数的参数。这个函数调用的结果是参数的类型。

我们通常说函数“接收”参数，并“返回”结果。这个结果称为返回值（return value）。

## 3.2 类型转换函数

Python提供了一些可将某个值从一种类型转换为另一种类型的内置函数。`int`函数可以把任何可以转换为整型的值转换为整型；如果转换失败，则会报错：

```
>>> int('32')
32
>>> int('Hello')
ValueError: invalid literal for int(): Hello
```

`int`可以将浮点数转换为整数，但不会做四舍五入操作，而是直接舍弃小数部分。

```
>>> int(3.99999)
3
>>> int(-2.3)
-2
```

`float`函数将整数和字符串转换为浮点数：

```
>>> float(32)
32.0
>>> float('3.14159')
3.14159
```

最后，`str`函数将参数转换为字符串：

```
>>> str(32)
'32'
>>> str(3.14159)
'3.14159'
```

### 3.3 数学函数

Python有一个数学计算模块，提供了大多数常用的数学函数。模块是指包含一组相关的函数的文件。

要想使用一个模块，需要先将它导入（import）运行环境：

```
>>> import math
```

这个语句将会建立一个名为math的模块对象（module object）。如果打印这个对象，可以看到它的一些信息：

```
>>> print math
<module 'math' (built-in)>
```

模块对象包含了这个模块中定义的函数和变量。若要访问其中的一个函数，需要同时指定模块名称和函数名称，用一个句点（.）分隔。这个格式称为句点表示法（dot notation）。

```
>>> ratio = signal_power / noise_power
>>> decibels = 10 * math.log10(ratio)

>>> radians = 0.7
>>> height = math.sin(radians)
```

上面第一个例子使用了log10来计算以分贝为单位的信号/噪声比（假设signal\_power和noise\_power都已经事先定义好了）。math模块也提供了log函数，用来计算底为e的自然对数。

第二个例子计算radians的正弦值。这个变量名已经暗示了，sin以及cos、tan等三角函数接受的参数是以弧度（radians）为单位的。若要将角度转换为弧度，可以除以 $360$ 再乘以 $2\pi$ ：

```
>>> degrees = 45
>>> radians = degrees / 360.0 * 2 * math.pi
>>> math.sin(radians)
0.707106781187
```

表达式math.pi从math模块中获得变量pi。这个变量的值是 $\pi$ 的近似值，大约精确到15位数字。

如果你了解三角函数，可以把上面的结果和 $\sqrt{2}$ 的一半进行比较：

```
>>> math.sqrt(2) / 2.0  
0.707106781187
```

## 3.4 组合

到现在为止，我们已经分别了解了程序的基本元素——变量、表达式和语句，但还没有接触如何将它们有机地组合起来。

程序设计语言最有用的特性之一就是可以将各种小的构建块（building block）组合起来。比如，函数的参数可以是任何类型的表达式，包括算术符号：

```
x = math.sin(degrees / 360.0 * 2 * math.pi)
```

甚至还包括函数调用：

```
x = math.exp(math.log(x+1))
```

基本上，在任何可以使用值的地方，都可以使用任意表达式，只有一个例外：赋值表达式的左边必须是变量名称，在左边放置任何其他的表达式都是语法错误（后面我们还会看到这条规则的例外情况）。

```
>>> minutes = hours * 60          # 正确
>>> hours * 60 = minutes         # 错误!
SyntaxError: can't assign to operator
```

## 3.5 添加新函数

至此，我们都只是在使用Python提供的函数，其实我们也可以自己添加新的函数。函数定义指定新函数的名称，并提供一系列程序语句。当函数被调用时，这些语句会顺序执行。

下面是一个例子：

```
def print_lyrics():
    print "I'm a lumberjack, and I'm okay."
    print "I sleep all night and I work all day."
```

`def`是关键字，表示接下来是一个函数定义。这个函数的名称是`print_lyrics`。函数名称的书写规则和变量名称一样：字母、数字和某些标点是合法的，但第一个字符不能是数字。关键字不能作为函数名，而且我们应尽量避免函数和变量同名。

函数名后的空括号表示它不接收任何参数。

函数定义的第一行称为函数头（header），其他部分称为函数体（body）。函数头应该以冒号结束，函数体整体缩进一级。依照惯例，缩进总是使用4个空格，参看3.14节。函数体的代码语句行数不限。

本例中`print`语句里的字符串使用双引号括起来。单引号和双引号的作用相同。大部分情况下，人们都使用单引号，只在本例中这样的特殊情况下才使用双引号。本例中的字符串里本身就存在单引号（单引号也作为省略符号用，如`I'm`）。

如果在交互模式里输入函数定义，则解释器会输出省略号（...）提示你当前的定义还没有结束：

```
>>> def print_lyrics():
...     print "I'm a lumberjack, and I'm okay."
...     print "I sleep all night and I work all day."
... 
```

想要结束这个函数的定义，需要输入一个空行（在脚本文件中则不需要如此）。

定义一个函数后，会创建一个同名的变量。

```
>>> print print_lyrics  
<function print_lyrics at 0xb7e99e9c>  
>>> type(print_lyrics)  
<type 'function'>
```

变量print\_lyrics的值是一个函数对象，其类型是'function'。

调用新创建的函数的方式，与调用内置函数是一样的：

```
>>> print_lyrics()  
I'm a lumberjack, and I'm okay.  
I sleep all night and I work all day.
```

定义好一个函数之后，就可以在其他函数中调用它。比如，若想重复上面的歌词，我们可以写一个repeat\_lyrics函数：

```
def repeat_lyrics():  
    print_lyrics()  
    print_lyrics()
```

然后可以调用repeat\_lyrics：

```
>>> repeat_lyrics()  
I'm a lumberjack, and I'm okay.  
I sleep all night and I work all day.  
I'm a lumberjack, and I'm okay.  
I sleep all night and I work all day.
```

当然，这首歌其实并不是这么唱的。

## 3.6 定义和使用

将前面一节零散的代码整合起来，整个程序就像下面这个样子：

```
def print_lyrics():
    print "I'm a lumberjack, and I'm okay."
    print "I sleep all night and I work all day."

def repeat_lyrics():
    print_lyrics()
    print_lyrics()

repeat_lyrics()
```

这个程序包含两个函数定义：print\_lyrics和repeat\_lyrics。（在解释器执行程序代码时）函数定义的执行方式和其他语句一样，不同的是执行后会创建函数对象。函数体里面的语句并不会立即执行，而是等到函数被调用时才执行。函数定义不会产生任何输出。

你可能已经猜到，必须先创建一个函数，才能执行它。换言之，函数定义必须在函数的第一次调用之前先执行。

### 练习3-1

将程序的最后一行移动到首行，于是函数调用会先于函数定义执行。运行程序并查看会有什么样的错误信息。

### 练习3-2

将函数调用那一行放回到末尾，并将函数print\_lyrics的定义放到函数repeat\_lyrics定义之后。这时候运行程序会发生什么？

## 3.7 执行流程

为了保证函数的定义先于其首次调用执行，你需要知道程序中语句执行的顺序，即**执行流程**。

执行总是从程序的第一行开始。从上到下，按顺序，每次执行一条语句。

函数定义并不会改变程序的执行流程，但应注意函数体中的语句并不立即执行，而是等到函数被调用时执行。

函数调用可以看作程序执行流程中的一个迂回路径。遇到函数调用时，并不会直接继续执行下一条语句，而是跳到函数体的第一行，继续执行完函数体的所有语句，再跳回到原来离开的地方。

这样看似简单，但马上你会发现，函数体中可以调用其他函数。当程序流程运行到一个函数之中时，可能需要执行其他函数中的语句。但当执行那个函数中的语句时，又可能再需要调用执行另一个函数的语句！

幸好Python对于它运行到哪里有很好的记录，所以每个函数执行结束后，程序都能跳回到它离开的地方。直到执行到整个程序的结尾，才会结束程序。

前面这段枯燥的描述，寓意何在？当你阅读代码时，并不总是应该一行行按照书写顺序阅读。有时候，按照执行的流程来阅读代码，可能理解效果更好。

## 3.8 形参和实参①

我们已经看到，有些内置函数需要传入参数。比如，当调用`math.sin`时，需要传入一个数字作为实参。有的函数需要多个实参：`math.pow`需要两个，分别是基数（`base`）和指数（`exponent`）。

在函数内部，实参会被赋值给形参。下面的例子是一个用户自定义的函数，接收一个实参：

```
def print_twice(bruce):
    print bruce
    print bruce
```

这个函数在调用时会把实参的值赋到形参`bruce`上，并将其打印两次。

这个函数对任何可以打印的值都可用。

```
>>> print_twice('Spam')
Spam
Spam
>>> print_twice(17)
17
17
>>> print_twice(math.pi)
3.14159265359
3.14159265359
```

内置函数的组合规则，在用户自定义函数上也同样可用，所以我们可以对`print_twice`使用任何表达式作为实参：

```
>>> print_twice('Spam '*4)
Spam Spam Spam Spam
Spam Spam Spam Spam
>>> print_twice(math.cos(math.pi))
-1.0
-1.0
```

作为实参的表达式会在函数调用之前先执行。所以在这个例子中，表达式`'Spam' * 4`和`math.cos(math.pi)`都只执行一次。

你也可以使用变量作为实参：

```
>>> michael = 'Eric, the half a bee.'  
>>> print_twice(michael)  
Eric, the half a bee.  
Eric, the half a bee.
```

作为实参传入到函数的变量的名称（michael）和函数定义里形参的名称（bruce）没有关系。函数内部只关心形参的值，而不用关心它在调用前叫什么名字；在print\_twice函数内部，大家都叫bruce。

## 3.9 变量和形参是局部的

当你在函数体内新建一个变量时，它是局部的（local），即它只存在于这个函数之内。比如：

```
def cat_twice(part1, part2):
    cat = part1 + part2
    print_twice(cat)
```

这个函数接收两个实参，将它们拼接起来，并将结果打印两遍。下面是一个使用这一函数的例子：

```
>>> line1 = 'Bing tiddle '
>>> line2 = 'tiddle bang.'
>>> cat_twice(line1, line2)
Bing tiddle tiddle bang.
Bing tiddle tiddle bang.
```

当cat\_twice结束时，变量cat会被销毁。这时再尝试打印它的话，会得到一个异常：

```
>>> print cat
NameError: name 'cat' is not defined
```

形参也是局部的。比如，在print\_twice函数之外，不存在bruce这个变量。

## 3.10 栈图

要跟踪哪些变量在哪些地方使用，有时候画一个栈图（stack diagram）会很方便。和状态图一样，栈图可以展示每个变量的值，不同的是它会展示每个变量所属的函数。

每个函数使用一个帧包含，帧在栈图中就是一个带着函数名称的盒子，里面有函数的参数和变量。前面的函数示例的栈图如图3-1所示。

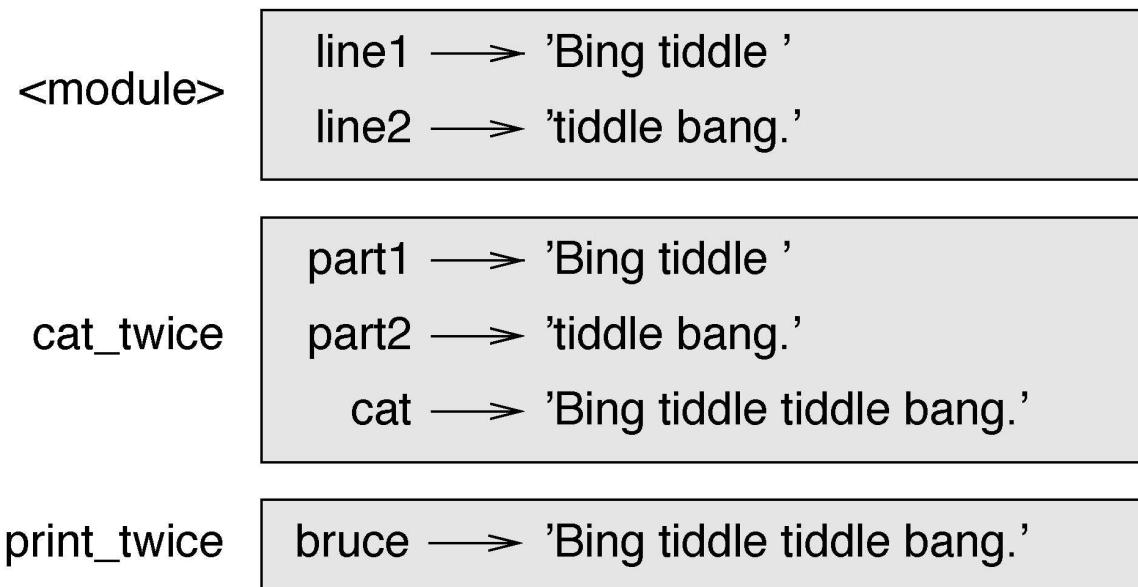


图3-1 栈图

图中各个帧从上到下安排成一个栈，能够展示出哪个函数被哪个函数调用了。在这个例子里，`print_twice`被`cat_twice`调用，而`cat_twice`被`__main__`调用。`__main__`是用于表示整个栈图的图框的特别名称。当你在所有函数之外新建变量时，它就是属于`__main__`的。

每个形参都指向与其对应的实参相同的值，所以，`part1`和`line1`的值相同，`part2`和`line2`的值相同，而`bruce`和`cat`的值相同。

如果调用函数的过程中发生了错误，Python会打印出函数名，以及调用它的函数的名称，以及调用这个调用者的函数名，依此类推，一直到`__main__`。

比如，如果你在print\_twice中访问cat变量，则会得到一个NameError：

```
Traceback (innermost last):
  File "test.py", line 13, in __main__
    cat_twice(line1, line2)
  File "test.py", line 5, in cat_twice
    print_twice(cat)
  File "test.py", line 9, in print_twice
    print cat
NameError: name 'cat' is not defined
```

上面这个函数列表被称为回溯（traceback）。它告诉你错误出现在哪个程序文件，哪一行，以及哪些函数正在运行。它也会显示导致错误的那一行代码。

回溯中函数的顺序和栈图中图框的顺序一致。当前正在执行的函数列在最底部。

## 3.11 有返回值函数和无返回值函数

我们使用过的函数中，有一部分函数，如数学函数，会产生结果。因为没有想到更好的名字，我称这类函数为有返回值函数（fruitful function）。另一些函数，如`print_twice`，会执行一个动作，但不返回任何值。我们称这类函数为无返回值函数（void function）。

当你调用一个有返回值的函数时，大部分情况下都想要对结果做某种操作。比如，你可能会想把它赋值给一个变量，或者用在一个表达式中：

```
x = math.cos(radians)
golden = (math.sqrt(5) + 1) / 2
```

在交互模式中调用函数时，Python会直接显示结果：

```
>>> math.sqrt(5)
2.2360679774997898
```

但是在脚本中，如果只是直接调用这类函数，那么它的返回值就会永远丢失掉！

```
math.sqrt(5)
```

这个脚本计算5的平方根，但由于并没有把计算结果存储到某个变量中，或显示出来，所以其实没什么实际作用。

无返回值函数可能在屏幕上显示某些东西，或者有其他的效果，但是它们没有返回值。如果你试着把它们的结果赋值给某个变量，则会得到一个特殊的值`None`。

```
>>> result = print_twice('Bing')
Bing
Bing
>>> print result
None
```

值`None`和字符串`'None'`并不一样。它是一个特殊的值，有自己独特的类型：

```
>>> print type(None)
<type 'NoneType'>
```

到目前为止，我们自定义的函数都是无返回值函数。再过几章我们就会开始写有返回值的函数了。

## 3.12 为什么要有函数

为什么要花功夫将程序拆分成函数呢？也许刚开始编程的时候这其中的原因并不明晰。下面这些解释都可作为参考。

- 新建一个函数，可以让你有机会给一组语句命名，这样可以让代码更易读和更易调试。
- 函数可以通过减少重复代码使程序更短小。后面如果你需要修改代码，也只要修改一个地方即可。
- 将一段程序拆分成几个函数后，可以对每一个函数单独进行调试，再将它们组装起来成为完整的产品。
- 一个设计良好的函数，可以在很多程序中使用。书写一次，调试一次，复用无穷。

## 3.13 使用from导入模块

Python提供了两种导入模块的方式；我们已经见过其中一种：

```
>>> import math
>>> print math
<module 'math' (built-in)>
>>> print math.pi
3.14159265359
```

如果你导入math，则会得到名为math的模块对象。模块对象包含了pi这样的常量以及诸如sin和exp这样的函数。

但是如果直接访问pi，则会发生错误。

```
>>> print pi
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
NameError: name 'pi' is not defined
```

这时候，你可以像下面这样来导入模块中的某个对象：

```
>>> from math import pi
```

现在就可以直接访问pi，而不需要使用句点表示法math.pi了。

```
>>> print pi
3.14159265359
```

或者，也可以使用星号来导入一个模块的所有成员：

```
>>> from math import *
>>> cos(pi)
-1.0
```

用这种方式导入模块内所有的成员，好处是可以使你的代码更简洁，但缺点是不同模块的同名成员之间，或者和自定义的变量之间，可能发生名字冲突。

## 3.14 调试

如果你使用文本编辑器来编写脚本，则可能会遇到缩进时空格和制表符混淆的问题。避免这种问题的最好办法是只使用空格（不用制表符）。大部分识别Python的文本编辑器都默认这么处理，不过也有一些不支持。

制表符和空格都是不可见的，因而会很难调试，所以应尝试找一个能帮你处理缩进的编辑器。

另外，不要忘了在运行程序前保存它。有的开发环境会自动保存，但也有不自动保存的。如果不保存，则你写好的代码和运行的代码并不一样。

如果运行的报错的代码和你写的不一样，调试时会浪费很多时间！

所以一定要确保你眼前所看的代码和所运行的代码是一致的。如果不確定，可以在程序开头写一句`print 'hello'`并再运行一次。如果没有看到hello输出，则你运行的不是正确的程序！

## 3.15 术语表

函数 (function)：一个有名称的语句序列，可以进行某种有用的操作。函数可以接收或者不接收参数，可以返回或不返回结果。

函数定义 (function definition)：一个用来创建新函数的语句，指定函数的名称、参数以及它执行的语句序列。

函数对象 (function object)：函数定义所创建的值。函数名可以用作变量来引用一个函数对象。

函数头 (header)：函数定义的第一行。

函数体 (body)：函数定义内的语句序列。

形参 (parameter)：函数内使用的用来引用作为实参传入的值的名称。

函数调用 (function call)：执行一个函数的语句。它由函数名称和参数列表组成。

实参 (argument)：当函数调用时，提供给它的值。这个值会被赋值给对应的形参。

局部变量 (local variable)：函数内定义的变量。局部变量只能在函数体内使用。

返回值 (return value)：函数的结果。如果函数被当做表达式调用，返回值就是表达式的值。

有返回值函数 (fruitful function)：返回一个值的函数。

无返回值函数 (void function)：没有返回值的函数。

模块 (module)：一个包含相关函数以及其他定义的集合的文件。

import语句 (import statement)：读入一个模块文件，并创建一个模块对象的语句。

**模块对象 (module object)** : 使用 `import` 语句时创建的对象，提供对模块中定义的值的访问。

**句点表示法 (dot notation)** : 调用另一个模块中的函数的语法，使用模块名加上一个句点符号，再加上函数名。

**组合 (composition)** : 使用一个表达式作为更大的表达式的一部分，或者使用语句作为更大的语句的一部分。

**执行流程 (flow of execution)** : 程序运行中语句执行的顺序。

**栈图 (stack diagram)** : 函数栈的图形表达形式，也展示它们的变量，以及这些变量引用的值。

**图框 (frame)** : 栈图中的一个图框，表达一个函数调用。它包含了局部变量以及函数的参数。

**回溯 (traceback)** : 当异常发生时，打印出正在执行的函数栈。

## 3.16 练习

### 练习3-3

Python提供了一个内置函数len，返回一个字符串的长度。所以len('allen')的值是5。

编写一个函数right\_justify，接收一个字符串形参s，并打印出足够的前导空白，以达到最后一个字符显示在第70列上。

```
>>> right_justify('allen')
                    allen
```

### 练习3-4

函数对象是一个值，你可以将它赋值给变量，或者作为实参传递。例如do\_twice是一个函数，接收一个函数对象作为实参，并调用它两次：

```
def do_twice(f):
    f()
    f()
```

下面是一个使用do\_twice来调用一个print\_spam函数两次的示例：

```
def print_spam():
    print 'spam'

do_twice(print_spam)
```

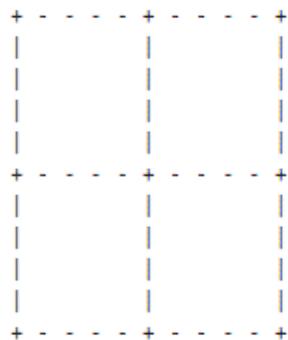
1. 将这个示例存入脚本中并测试它。
2. 修改do\_twice，让它接收两个实参，一个是函数对象，另一个是一个值，它会调用函数对象两次，并传入那个值作为实参。
3. 编写一个更通用的print\_spam，叫做print\_twice，接收一个字符串形参，并打印它两次。
4. 使用修改版的do\_twice来调用print\_twice两次，并传入实参'spam'。

5. 定义一个新的函数`do_four`，接收一个函数对象与一个值，使用这个值作为实参调用函数4次。这个函数的函数体应该只有两个语句，而不是四个。

### 练习3-5

这个练习可以只用语句和我们已经学过的其他语言特性实现。

1. 编写一个函数，绘制如下的表格：



#### 提示：

要在同一行打印多个值，你可以使用逗号分隔不同的值：

```
print '+', '-'
```

如果值序列的结尾有一个逗号，Python不会换行，所以后面的打印语句会出现在同一行。

```
print '+',  
print '-'
```

这两个语句的输出是'+ -'。

2. 编写一个函数绘制类似的表格，但有4行4列。

解答：<http://thinkpython.com/code/grid.py>。

鸣谢：这个练习基于Oualline的《实践C编程》第3版（O'Reilly Media, 1997）中的一个示例。

---

① 这一段中讲的参数有两种：函数定义里的形参（parameter），以及调用函数时传入的实参（argument），这里两种是有区分的。——译者注



# 第4章 案例研究：接口设计

本章的代码示例可以从<http://thinkpython.com/code/polygon.py>下载。

## 4.1 乌龟世界

为了配合本书，我写了一个程序包，称为Swampy。你可以从<http://thinkpython.com/swampy>下载Swampy；参考网站上的指导，将Swampy安装到你的系统中。

程序包（package）是多个模块的组合；Swampy中有一个模块“乌龟世界”（TurtleWorld），它提供各种函数，可以引导一只乌龟在屏幕上爬行，并画出其踪迹。

系统中安装好了Swampy之后，就可以像下面这样导入TurtleWorld模块：

```
from swampy.TurtleWorld import *
```

如果你下载了Swampy但并没有安装，则可以在其代码目录中使用，或者将其目录加入到Python的搜索路径中。接下来就可以这样导入TurtleWorld：

```
from TurtleWorld import *
```

安装过程的细节和如何设置Python的搜索路径，依赖于你所使用的系统。所以我在这里不细加讨论，而在<http://thinkpython.com/swampy>里维护几种系统的相关安装信息。

创建一个文件mypolygon.py，并输入如下代码：

```
from swampy.TurtleWorld import *

world = TurtleWorld()
bob = Turtle()
print bob

wait_for_user()
```

第一行从swampy程序包的TurtleWorld模块里导入全部成员。

接下来几行建立一个TurtleWorld对象，赋值给变量world；建立一个Turtle对象，赋值给bob。

打印bob对象会得到如下信息：

```
<TurtleWorld.Turtle instance at 0xb7bfbf4c>
```

这意味着bob变量引用着TurtleWorld模块中定义的Turtle类的一个实例（instance）。在这个语境中，“实例”是指集合中的一员；bob这个Turtle对象是可能存在的所有Turtle的集合的一员。

wait\_for\_user告诉TurtleWorld等待用户进行某些操作，虽然现在除了关闭窗口之外，并没有提供给用户多少有用的操作。

TurtleWorld提供了几个用来指挥乌龟的函数：fd和bk用于前进和后退，lt和rt用于左转和右转。另外，每只乌龟都拿着一只笔，可以朝上或者朝下；若笔朝下，则会绘制出走过的路迹。函数pu和pd分别表示“笔朝上”（pen up）和“笔朝下”（pen down）。

若要画一个朝右的角，在程序中（建立bob实例之后，调用wait\_for\_user之前）添加如下代码：

```
fd(bob, 100)
lt(bob)
fd(bob, 100)
```

第一行告诉bob前进100步。第二行告诉它左拐。

当你运行这个程序时，将会看到bob先向东走，再向北走，身后留下两条线段。

现在试着修改程序，画出一个正方形来。在成功之前请不要继续！

## 4.2 简单重复

你可能会写下如下代码（除去新建TurtleWorld和等待用户的操作外）：

```
fd(bob, 100)
lt(bob)

fd(bob, 100)
lt(bob)

fd(bob, 100)
lt(bob)

fd(bob, 100)
```

使用`for`语句，可以更紧凑地实现同样功能。把下面的例子加到`mypolygon.py`中，并再运行一次：

```
for i in range(4):
    print 'Hello!'
```

你可能会看到如下输出：

```
Hello!
Hello!
Hello!
Hello!
```

这是`for`语句的最简单用法，后面我们会看到更多的用法。但这样已经足够重写刚才的画正方形的程序了。请重写后再接着阅读。

下面是使用`for`语句绘制正方形的程序：

```
for i in range(4):
    fd(bob, 100)
    lt(bob)
```

`for`语句的语法和函数定义类似。它也有一个以冒号结束的语句头，并有一个缩进的语句体。语句体可以包含任意数量的语句。

`for`语句有时被称为循环（loop），因为执行流程会遍历语句体，之后从语句体的最开头重新循环执行。在这个例子里，语句体执行了4次。

这个版本的代码和之前的绘制正方形的代码其实还稍有不同，因为在最后一次循环后它多做了一次左转。多余的左转稍微多消耗了点时间，但因为每次循环做的事情都一样，也让代码更简练。这个版本的代码还有一个效果，程序执行完之后，乌龟会回归到初始的位置，并朝向初始相同的方向。

## 4.3 练习

下面是一系列使用TurtleWorld的练习。它们力求有趣，但也包含着某些寓意。当你做这些练习时，可以猜想一下其寓意。

在接下来的章节中有这些练习的解答，所以在完成（或者至少尝试过）之前，请先别继续阅读。

1. 写一个函数square，接受一个形参t，用来表示一只乌龟。利用乌龟来画一个正方形。写一个函数调用传入bob作为实参来调用square函数，并再运行一遍程序。

2. 给square函数再添加一个形参length。修改函数内容，保证正方形的长度是length，并修改函数调用以提供这第二个实参。再运行一遍程序。使用不同的length值测试你的程序。

3. 函数lt和rt默认会进行90度的拐弯。但也可以提供第二个形参指定转弯的角度。比如，lt(bob, 45)会让bob左拐45度。

复制square函数，并命名为polygon。再添加一个形参n并修改函数体以绘制一个正n边形。提示：正n边形的拐角是 $360/n$ 度。

4. 写一个函数circle接受代表乌龟的形参t，以及表示半径的形参r，并使用合适的长度和边数调用polygon画一个近似的圆。使用不同的r值来测试你的函数。

### 提示：

思考圆的周长（circumference），并保证length \* n = circumference。

另一个提示：如果你感觉bob太慢了，可以修改bob.delay来加速。bob.delay代表每次行动之间的停顿，单位是秒。bob.delay = 0.01应该能让它跑得足够快。

5. 给circle函数写一个更通用的版本，称为arc。增加一个形参angle，用来表示画的圆弧的大小。这里angle的单位是度数，所以当arc=360时，则会画一个整圆。

## 4.4 封装

第一个练习要求你把画正方形的代码放到一个函数定义中，并将乌龟bob作为实参传入，调用该函数。下面是一个解答：

```
def square(t):
    for i in range(4):
        fd(t, 100)
        lt(t)

square(bob)
```

最内侧的语句，`fd`和`lt`都缩进了两层，表示它们是在`for`语句的语句体内部，而`for`语句在函数定义的函数体内部。最后一行，`square(bob)`，又重新从左侧开始而没有缩进，所以这里`for`语句和`square`函数的定义都已经结束了。

在函数体中，`t`引用的乌龟和`bob`引用的相同，所以`lt(t)`和直接调用`lt(bob)`是一样的效果。那么为什么不直接把形参写为`bob`呢？原因是`t`可以是任何乌龟，而不仅仅是`bob`，所以你可以再新建一只乌龟，并将它作为参数传入到`square`函数：

```
ray = Turtle()
square(ray)
```

把一段代码用函数包裹起来，称为封装（encapsulation）。封装的一个好处是，它给这段代码一个有意义的名称，增加了可读性。另一个好处是，当你重复使用这段代码时，调用一次函数比复制粘贴代码要简易得多！

## 4.5 泛化

下一步是给square函数添加一个length参数。这里是一个解决方案：

```
def square(t, length):
    for i in range(4):
        fd(t, length)
        lt(t)

square(bob, 100)
```

给函数添加参数的过程称为泛化（generalization），因为它会让函数变得更通用：在之前的版本中，正方形总是一个大小，而新的版本中，可以是任意大小。

下一步也是一次泛化。我们不再只绘制正方形，而是可以绘制任意边数的多边形。这里是一个方案：

```
def polygon(t, n, length):
    angle = 360.0 / n
    for i in range(n):
        fd(t, length)
        lt(t, angle)

polygon(bob, 7, 70)
```

这个例子绘制一个7边形，边长是70。如果函数的形参比较多，很容易忘掉每一个具体是什么，或者忘掉它们的顺序。所以在Python中，调用函数时可以加上形参名称，这样是合法的，并且有时候会有帮助：

```
polygon(bob, n=7, length=70)
```

这些参数被称为关键词参数（keyword argument），因为它们使用“关键词”的形式带上了形参的名称调用（请别和while与def之类的Python关键字混淆）。

这个语法使得程序更加可读。它也同样提示了我们实参和形参的工作方式：当调用函数时，实参传入并赋值给形参。

## 4.6 接口设计

下一步是写画圆的circle函数，接受形参r，表示圆的半径。下面是一个简单的例子，通过调用polygon函数画50边的多边形：

```
def circle(t, r):
    circumference = 2 * math.pi * r
    n = 50
    length = circumference / n
    polygon(t, n, length)
```

第一行计算半径为r的圆的周长，使用公式 $2\pi r$ 。因为我们使用的是math.pi，所以需要先导入math模块。依照惯例，import语句一般都放在脚本开头。

n是我们用于近似画圆的多边形的边数，所以length是每个边的长度。因此，polygon画出一个50边形，近似于一个半径为r的圆。

这个解决方案的缺点之一是n是一个常量，因此对于很大的圆，多边形的边线太长，而对于小圆，我们又浪费时间去画过短的边线。解决办法之一是泛化这个函数，加上形参n。这样可以给用户（调用circle函数的人）更多的控制选择，但接口就不那么清晰简洁了。

函数的接口是如何使用它的概要说明：它有哪些参数？这个函数做什么？它的返回值是什么？我们说一个接口“简洁”（clean），是希望它“尽可能简单，但不能过度。（爱因斯坦）”。

在这个例子里，r属于函数的接口，因为它指定了所画的圆的基本属性。相对地，n则不那么适合，因为它说明的是如何画圆的细节信息。

所以与其弄乱接口，不如在代码内部根据周长来选择合适的n值：

```
def circle(t, r):
    circumference = 2 * math.pi * r
    n = int(circumference / 3) + 1
    length = circumference / n
    polygon(t, n, length)
```

现在多边形的边数（近似）是circumference/3，所以每个边长（近似）是3，已经小到足够画出好看的圆形，但又足够大到不影响画线效

率，并且适合于任何尺寸的圆。

## 4.7 重构

当我写circle函数时，我可以复用polygon，因为边数很多的正多边形是圆的很好的近似。但是arc则并不那么容易对付；我们不能使用polygon或者circle来画圆弧。

一个选择是先复制一个polygon函数，再通过修改得到arc函数。结果可能类似下面的示例：

```
def arc(t, r, angle):
    arc_length = 2 * math.pi * r * angle / 360
    n = int(arc_length / 3) + 1
    step_length = arc_length / n
    step_angle = float(angle) / n

    for i in range(n):
        fd(t, step_length)
        lt(t, step_angle)
```

这个函数的第二部分很像polygon的实现，但如果修改polygon的接口，无法直接复用。我们也可以泛化polygon函数以接受第三个参数表示圆弧的角度，但那样的话polygon（多边形）就不是合适的名称了！所以，我们将这个更泛化的函数称为polyline（多边线）：

```
def polyline(t, n, length, angle):
    for i in range(n):
        fd(t, length)
        lt(t, angle)
```

现在我们可以重写polygon和arc，让它们调用polyline：

```
def polygon(t, n, length):
    angle = 360.0 / n
    polyline(t, n, length, angle)

def arc(t, r, angle):
    arc_length = 2 * math.pi * r * angle / 360
    n = int(arc_length / 3) + 1
    step_length = arc_length / n
    step_angle = float(angle) / n
    polyline(t, n, step_length, step_angle)
```

最后，我们可以重写circle，改为调用arc：

```
def circle(t, r):
    arc(t, r, 360)
```

这个过程——重新组织程序，以改善函数的接口，提高代码复用——被称为重构（refactoring）。在这个例子里，我们注意到arc和polygon中有类似的代码，因此我们把它们的共同之处“重构出来”抽取到 polyline函数中。

如果我们早早计划，可能会直接先写下polyline，也就避免了重构，但实际上在工程开始时我们往往并没有足够的信息去完美设计所有的接口。开始编码之后，你会更了解面对的问题。有时候，重构正意味着你在编程中掌握了一些新的东西。

## 4.8 一个开发计划

开发计划 (development plan) 是写程序的过程。本章的案例分析中，我们使用的过程是“封装和泛化”。这个过程的具体步骤是：

1. 最开始写一些小程序，而不需要函数定义。
2. 一旦程序成功运行，将它封装到一个函数中，并加以命名。
3. 泛化这个函数，添加合适的形参。
4. 重复步骤1到步骤3，直到你得到一组可行的函数。复制粘贴代码，以避免重复输入（以及重复调试）。
5. 寻找可以使用重构来改善程序的机会。例如，如果发现程序中几处地方有相似的代码，可以考虑将它们抽取出来做一个合适的通用函数。

这个过程也有一些缺点——我们会在后面看到其他方式——但如果你在开始编程时不清楚如何将程序分成适合的函数，这样做会带来帮助。这个方法能让你一边开发一边设计。

## 4.9 文档字符串

文档字符串（docstring）是在函数开头用来解释其接口的字符串（doc是“文档” documentation的缩写）。下面是一个示例：

```
def polyline(t, n, length, angle):
    """Draws n line segments with the given length and
    angle (in degrees) between them. t is a turtle.
    """
    for i in range(n):
        fd(t, length)
        lt(t, angle)
```

这里的文档字符串是一个使用三引号括起来的字符串。三引号字符串又称为多行字符串，因为三引号允许字符串跨行表示。

文档字符串很简洁，但已经包含了其他人需要知道的关于函数的基本信息。它精确地解释了函数是做什么的（而不会涉及它是如何实现的细节）。它解释了每个形参对函数行为的影响效果以及每个形参应有的类型（如果其类型并不显而易见）。

编写这类文档是接口设计的重要部分。一个设计良好的接口，也应当很简单就能解释清楚；如果你发现解释一个函数很困难，很可能表示它的接口设计有改进的空间。

## 4.10 调试

函数的接口，作用就像是函数和调用者之间签订的一个合同。调用者同意提供某些参数，而函数则同意使用这些参数做某种工作。

例如，`polyline`需要4个参数：`t`必须是一个Turtle；`n`是边线的个数，所以必须是整数；`length`应当是个正数；而`angle`则必须是一个数字，并且按照度数来理解。

这些需求被称为前置条件，因为它们应当在函数开始执行之前就保证为真。相对地，函数结束的时候需要满足的条件称为后置条件。后置条件包含了函数预期的效果（比如画出线段）以及任何副作用（比如移动乌龟或者引起World中的其他改变）。

满足前置条件是调用者的职责。如果调用者违反了一个（文档说明清晰的！）前置条件，因而导致函数没有正确运行，则bug是在调用者，而不是在函数本身。

## 4.11 术语表

**实例 (instance) :** 一个集合中的一员。本章中的TurtleWorld是TurtleWorlds集合中的一员。

**循环 (loop) :** 程序中的一个片段，可以重复执行。

**封装 (encapsulation) :** 将一组语句转换为函数定义的过程。

**泛化 (generalization) :** 将一些不必要的具体值（如一个数字）替换为合适的通用参数或变量的过程。

**关键词参数 (keyword argument) :** 调用函数时，附带了参数名称（作为一个“关键词”来使用）的参数。

**接口 (interface) :** 描述函数如何使用的说明。包括函数的名称，以及形参与返回值的说明。

**重构 (refactoring) :** 修改代码并改善函数的接口以及代码质量的过程。

**开发计划 (development plan) :** 写程序的过程。

**文档字符串 (docstring) :** 在函数定义开始处出现的用于说明函数接口的字符串。

**前置条件 (precondition) :** 在函数调用开始前应当满足的条件。

**后置条件 (postcondition) :** 在函数调用结束后应当满足的条件。

## 4.12 练习

### 练习4-1

在<http://thinkpython.com/code/polygon.py>下载本章的代码。

1. 给函数polygon, arc和circle编写合适的文档字符串。
2. 画一个栈图来显示函数circle(bob, radius)运行时的程序状态。你可以手动计算，或者在代码中添加一些print语句。
3. 在4.7节中的arc函数并不准确，因为使用多边形模拟近似圆，总是在真实的圆之外。因此，乌龟画完线之后会停在偏离正确的目标几个单位的地方。我的解决方案里展示了一种方法可以减少这种错误的效果。阅读代码并考虑是否合理。如果你自己画图，可能会发现它是如何生效的。

### 练习4-2

写一组合适的通用函数，用来画出图4-1所示的花朵图案。

解答：<http://thinkpython.com/code/flower.py>，另外也需要<http://thinkpython.com/code/polygon.py>。

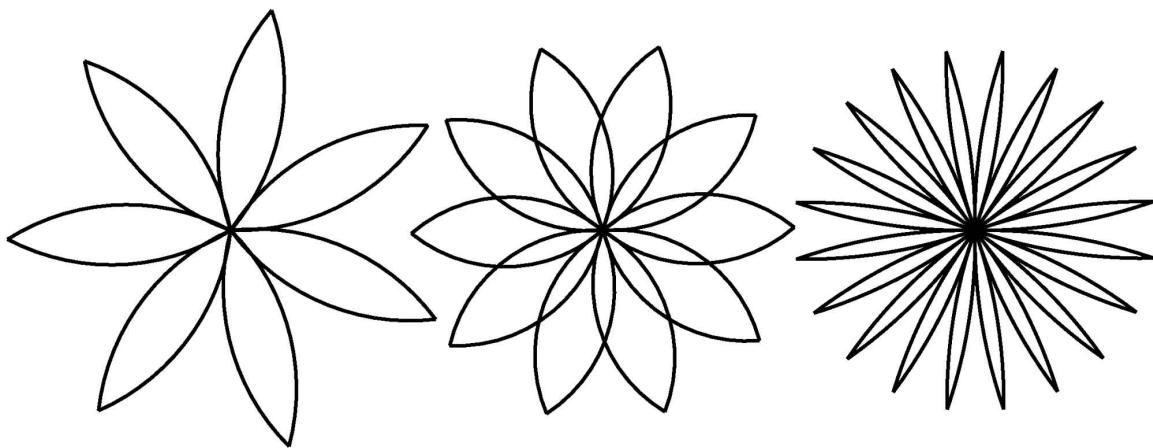


图4-1 花朵图案

### 练习4-3

写一组合适的通用函数，用来画出图4-2所示的图形。

解答：<http://thinkpython.com/code/pie.py>。

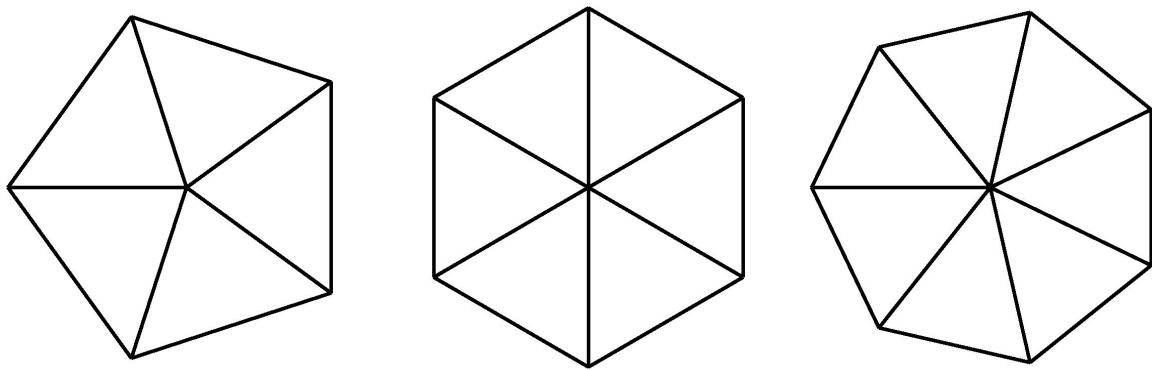


图4-2 饼图

#### 练习4-4

字母表中的字母可以使用一些基本元素来构成。如横线、竖线以及一些曲线。设计一个字体，可以使用最少的基本元素画出来，并编写函数来画出字母表中所有的字母。

你应当给每个字母单独写一个函数，名称为draw\_a、draw\_b等，并把这些函数放到letters.py文件中。可以从<http://thinkpython.com/code/typewriter.py> 下载一个“乌龟打字机”程序来帮助测试你的代码。

解答：<http://thinkpython.com/code/letters.py>，另外也需要<http://thinkpython.com/code/polygon.py>。

#### 练习4-5

在<http://en.wikipedia.org/wiki/Spiral>阅读关于螺旋线（spiral）的信息；接着编写一段程序来画出阿基米德螺旋（或者其他某种螺旋线）。

解答：<http://thinkpython.com/code/spiral.py>。

# 第5章 条件和递归

## 5.1 求模操作符

求模操作符作用于整数，可以计算出第一个操作数除以第二个操作数的余数。在Python中，求模操作符是一个百分号（%）。语法和其他的操作符一样：

```
>>> quotient = 7 / 3
>>> print quotient
2
>>> remainder = 7 % 3
>>> print remainder
1
```

所以7除以3的商是2，余数是1。

求模操作符有很多实际用途。比如，你可以用它来检测一个数是不是另一个的倍数——如果 $x \% y$ 是0，则x可以被y整除。

另外，你也可以用它来获取一个数后一位或后几位数字。例如， $x \% 10$ 可以得到x的个位数（10进制）。类似地， $x \% 100$ 可以获得最后两位数。

## 5.2 布尔表达式

布尔表达式是值为真或假的表达式。下面的例子中使用了`==`操作符，来比较两个操作对象是否相等。如果相等，则得`True`，否则是`False`:

```
>>> 5 == 5
True
>>> 5 == 6
False
```

`True`和`False`是类型`bool`的两个特殊值；它们不是字符串：

```
>>> type(True)
<type 'bool'>
>>> type(False)
<type 'bool'>
```

`==`操作符是一个关系操作符；其他的关系操作符有：

```
x != y          # x不等于y
x > y          # x比y大
x < y          # x比y小
x >= y         # x大于或等于y
x <= y         # x小于或等于y
```

虽然这些操作你可能已经熟悉，Python的符号和数学符号还是有些区别的。最常见的错误是使用单等号（`=`）而不是双等号（`==`）。请记住`=`是一个赋值操作符，而`==`是一个关系操作符。另外，不存在`=<`或者`=>`这样的操作符。

## 5.3 逻辑操作符

逻辑操作符有3个：`and`、`or`和`not`。这些操作符的语义（意义）和它们在英语中的意思差不多。比如，`x > 0 and x < 10`只有当`x`比0大且比10小时才为真。

`n%2 == 0 or n%3 == 0`，当其中任意一个条件为真时为真，也就是说，数`n`可以被2或3整除都可以。

最后，`not`操作符可以否定一个布尔表达式，所以`not (x > y)`在`x > y`为假时为真，即当`x`小于等于`y`时真。

严格地说，逻辑操作符的操作对象应该都是布尔表达式，但是Python并不那么严格。任何非0的数都被解释为“true”。

```
>>> 17 and True  
True
```

这种灵活性可能会很有用，但有时候也会带来一些小困惑。你可能应该避免使用它（除非你很确切地知道你在做什么）。

## 5.4 条件执行

为了编写有用的程序，我们几乎总是需要检查条件并据此改变程序的行为的能力。条件语句给了我们这种能力。最简单的形式是if表达式：

```
if x > 0:  
    print 'x is positive'
```

if之后的布尔表达式被称为条件（condition）。如果它为真，则之后缩进的语句会被执行。否则，什么都不发生。

if表达式的结构和函数定义一样：一个语句头，接着是缩进的语句体。这种类型的语句称为复合语句。

语句体中出现的语句数量并没有限制，但是最少需要一行。偶尔可能会遇到需要一个语句体什么都不做（通常是标记一个你还没有来得及写的代码的位置）。这个时候，可以使用pass语句。pass语句什么都不做。

```
if x < 0:  
    pass          # 需要处理负值的情况！
```

## 5.5 选择执行

`if`语句的第二种形式是选择执行，这种形式下，有两种可能，而`if`的条件决定哪一种可能被真正执行。语法看起来是这样的：

```
if x%2 == 0:  
    print 'x is even'  
else:  
    print 'x is odd'
```

如果`x`除以2的余数是0，则我们知道`x`是偶数（even），并且程序会依次显示出`'x is even'`。如果条件为假，则第二段语句会被执行。因为条件必定是真假之一，所以必然只会有一段语句被执行。这两段不同的语句称为分支（branch），因为它们是程序执行流程中的两个支流。

## 5.6 条件链

有时候有超过两种的可能，所以我们需要更多的分支。表达这种计算的一种方式是条件链（chained conditional）：

```
if x < y:  
    print 'x is less than y'  
elif x > y:  
    print 'x is greater than y'  
else:  
    print 'x and y are equal'
```

`elif`是“`else if`”的缩写。和之前一样，只有一个分支会被真正执行。`elif`语句的数量没有限制。如果有一个`else`语句，则它必须放在最后。但也可以没有`else`语句。

```
if choice == 'a':  
    draw_a()  
elif choice == 'b':  
    draw_b()  
elif choice == 'c':  
    draw_c()
```

每个条件都按顺序检查。如果第一个是`false`，则检查下一个，依此类推。如果有一个条件为真，则执行相应的分支，而整个语句结束。即使有多个条件为`true`，也只有第一个为真的分支会被执行。

## 5.7 嵌套条件

条件判断可以再嵌套条件判断。我们可以修改前面的三分法示例，如下：

```
if x == y:  
    print 'x and y are equal'  
else:  
    if x < y:  
        print 'x is less than y'  
    else:  
        print 'x is greater than y'
```

外侧的条件语句包含两个分支。第一个分支包含一行简单的语句。第二个分支则包含了另一个if语句，它本身也有两个分支。这两个分支也都是简单语句，虽然它们其实也可以是条件语句。

虽然语句的缩进让结构非常明晰，但嵌套条件语句会很快随着嵌套层数增多而变得非常难以阅读。通常来说，你应该尽量避免它。

逻辑操作符常常能够用来简化嵌套条件语句。例如，我们可以将下面的语句替换为单独的一个条件：

```
if 0 < x:  
    if x < 10:  
        print 'x is a positive single-digit number.'
```

print语句只有在两个条件语句都通过时才执行，所以我们可以使用and操作符达到相同的效果：

```
if 0 < x and x < 10:  
    print 'x is a positive single-digit number.'
```

## 5.8 递归

函数调用另外一个函数是合法的；函数调用自己也是合法的。这样做有什么好处可能还不明显，但它其实是程序能做的最神奇的事情之一。例如，考虑下面的函数：

```
def countdown(n):
    if n <= 0:
        print 'Blastoff!'
    else:
        print n
        countdown(n - 1)
```

如果n是0或负数，它会输出单词“Blastoff!”，其他情况下，它会输出n，并调用一个名为countdown的函数——它自己——并传入实参n-1。

我们调用这个函数时会发生什么？

```
>>> countdown(3)
```

countdown的执行从n=3开始，因为n比0大，所以会输出3，并接着调用自己……

countdown的执行从n=2开始，因为n比0大，所以会输出2，并接着调用自己……

countdown的执行从n=1开始，因为n比0大，所以会输出1，并接着调用自己……

countdown的执行从n=0开始，因为n不比0大，所以会输出单词“Blastoff!”，并返回。

接收n=1的函数countdown返回。

接收n=2的函数countdown返回。

接收n=3的函数countdown返回。

然后你就会到了\_\_main\_\_函数。所以，全部的输出如下：

```
3  
2  
1  
Blastoff!
```

调用自己的函数称为递归的（recursive）函数；这个过程叫做递归（recursion）。

另外举一个例子，我们可以写一个函数打印某个字符串n次。

```
def print_n(s, n):  
    if n <= 0:  
        return  
    print s  
    print_n(s, n-1)
```

如果 $n \leq 0$ ，`return`语句会直接退出当前函数。执行流程会立即返回到调用者，之后的语句不会执行。

函数另外的部分和`countdown`类似：如果 $n$ 大于0，它会打印 $s$ 并且调用自己，以再进行 $n-1$ 次显示 $s$ 的操作。所以输出的行数是 $1 + (n-1)$ ，也就是 $n$ 。

对于这样简单的例子来说，可能使用`for`循环会更容易。但我们在后面见到一些示例，使用`for`循环很难写，但使用递归则会很简单，所以早早开始了解递归是件好事。

## 5.9 递归函数的栈图

在3.10节中，我们使用一个栈图来表示程序在进行函数调用时的状态。同样的栈图，可以用来帮助我们解释递归函数。

一个函数每次被调用时，Python会新建一个函数帧（function frame），内含函数的局部变量和参数。对于递归函数，栈上可能同时存在多个函数帧。

图5-1展示了countdown函数在n=3调用时的栈图。

<module>

countdown

countdown

countdown

countdown

$n \rightarrow 3$

$n \rightarrow 2$

$n \rightarrow 1$

$n \rightarrow 0$

图5-1 栈图

和往常一样，栈的顶端是 `main` 的函数帧。因为我们没有在 `main` 函数里新建任何变量或传入任何参数，所以它是空的。

4个`countdown`函数帧有不同的参数n值。最底端的栈，其 $n=0$ ，被称为基准情形（base case）。因为它不再进行递归调用，所以后面没有其他函数帧了。

练习5-1

为函数print\_n画一个栈图，其调用实参是s = 'Hello'和n=2。

练习5-2

写一个函数do\_n，接受一个函数对象和一个数字n作为形参。它会调用给定的函数n次。

## 5.10 无限递归

如果一个递归永远达不到基准情形，则它会永远继续递归调用，而程序也永不停止。这个现象被称为无限递归，而它并不是个好主意。下面是一个会引起无限递归的最简单函数：

```
def recurse():
    recurse()
```

在大多数程序环境中，无限递归的函数并不会真的永远执行。Python会在递归深度到达上限时报告一个错误信息：

```
File "<stdin>", line 2, in recurse
File "<stdin>", line 2, in recurse
File "<stdin>", line 2, in recurse
...
.
.
.
File "<stdin>", line 2, in recurse
RuntimeError: Maximum recursion depth exceeded
```

这个调用回溯比上一章看到的要大一些。当这个错误发生时，栈上已经有1000个帧了！

## 5.11 键盘输入

目前为止我们写过的程序都还有些粗鲁，因为它们还不能接收用户的输入。它们只能每次做相同的事情。

Python 2提供了一个内置函数`raw_input`来从键盘获取输入。在Python 3里，这个函数叫做`input`。当这个函数被调用时，程序会停止运行，并等待用户输入一些东西。当用户按下回车键，程序会恢复运行，而且`raw_input`则通过字符串形式返回用户输入的内容。

```
>>> input = raw_input()
What are you waiting for?
>>> print input
What are you waiting for?
```

在从用户那里获得输入之前，最好打印一个提示信息，告诉用户希望他们输入什么。`raw_input`函数可以接受一个参数作为提示：

```
>>> name = raw_input('What...is your name?\n')
What...is your name?
Authur, King of the Britons!
>>> print name
Authur, King of the Britons!
```

提示信息最后的`\n`表示一个换行符，它是会引起输出显示换行的特殊字符。这也是为何用户的输入显示在提示信息的下一行的原因。

如果你希望用户输入一个整数，可以尝试将输入值转换为`int`：

```
>>> prompt = 'What...is the airspeed velocity of an unladen swallow?\n'
>>> speed = raw_input(prompt)
What...is the airspeed velocity of an unladen swallow?
17
>>> int(speed)
17
```

但如果用户输入不是数字的话，会得到错误：

```
>>> speed = raw_input(prompt)
What...is the airspeed velocity of an unladen swallow?
What do you mean, an African or a European swallow?
```

```
|>>> int(speed)
|ValueError: invalid literal for int()
```

后面我们会看到如何处理这种错误。

## 5.12 调试

当发生错误时，Python显示的回溯包含了大量的信息，但有时候反而会信息过量，尤其是当栈上有很多帧的时候。最有用的信息是：

- 错误的类型；
- 发生错误的地方。

语法错误通常都很容易定位，但也有棘手之处。空格问题引起的错误很难处理，因为空格和制表符都是不可见的，我们已经习惯于忽视它们。

```
>>> x = 5
>>> y = 6
  File "<stdin>", line 1
    y = 6
    ^
SyntaxError: invalid syntax
```

这个例子中，问题的原因是第二行多缩进了一个空格。但错误信息指向的是y，容易误导。总的来说，错误信息会告诉我们发现错误的地址，但真正发生的地方可能在更前面的代码中，有时候甚至在前一行。

运行时错误也是如此。

假设你想要按照分贝来计算信噪比。公式是：  
 $SNR_{db} = 10\lg(P_{signal}/P_{noise})$ 。在Python中，你可能会这么写：

```
import math
signal_power = 9
noise_power = 10
ratio = signal_power / noise_power
decibels = 10 * math.log10(ratio)
print decibels
```

但当在Python 2中运行这个程序时，会得到错误信息：

```
Traceback (most recent call last):
  File "snr.py", line 5, in ?
    decibels = 10 * math.log10(ratio)
OverflowError: math range error
```

错误信息指向第5行，但那一行其实没有什么错误。要找到真正的错误，可能需要打印出ratio的值，结果你会发现是0。问题出在第4行，因为两个整数相除是会进行舍去操作的。解决方案是使用浮点数来表示信号强度和噪声强度。

总的来说，错误信息告诉你发现问题的地方，但那常常并不是问题发生的地方。

在Python 3中，这个例子并不会引起错误；Python 3中，即使是整数，除法也会进行浮点数除法。

## 5.13 术语表

求模操作符 (modulus operator)：用%表示的操作符，用于两个整数，计算其相除的余数。

布尔表达式 (boolean expression)：一种表达式，其值是True或False。

关系操作符 (relational operator)：用来表示两个操作对象的比较关系的操作符，如下之一：==，!=，>，<，>=，以及<=。

逻辑操作符 (logical operator)：用来组合两个布尔表达式的操作符，有3个——and、or和not。

条件语句 (conditional statement)：依照某些条件控制程序执行流程的语句。

条件 (condition)：条件语句中的布尔表达式，由它决定执行哪一个分支。

复合语句 (compound statement)：一个包含语句头和语句体的语句。语句头以冒号 (:) 结尾。语句体相对语句头缩进一层。

分支 (branch)：条件语句中的一个可能性分支语句段。

条件链语句 (chained conditional)：一种包含多个分支的条件语句。

嵌套条件语句 (nested conditional)：在其他条件语句的分支中出现的条件语句。

递归 (recursion)：在当前函数中调用自己的过程。

基准情形 (base case)：递归函数中的一个条件分支，里面不会再继续递归调用。

无限递归 (infinite recursion)：没有基准情形的递归，或者永远无法达到基准情形的分支的递归调用。最终，这种无限递归会导致运行时错误。

## 5.14 练习

### 练习5-3

费马大定理是说对于任何大于2的 $n$ ，不存在任何整数 $a$ 、 $b$ 和 $c$ 能够满足：

$$a^n + b^n = c^n$$

1. 编写一个函数`check_fermat`，接收4个形参（即 $a$ 、 $b$ 、 $c$ 和 $n$ ）并检查费马定理是否成立。如果 $n$ 比2大并且满足

$$a^n + b^n = c^n$$

则程序应当打印“天哪，费马弄错了！”，否则程序应当打印“不，那样不行”。

2. 编写一个函数，提示用户输入 $a$ 、 $b$ 、 $c$ 和 $n$ 的值，将它们转换为整数，并使用`check_fermat`来检查它们是否违背了费马定理。

### 练习5-4

如果给你3根木棍，你可能可以将它们摆成一个三角形，也可能不可以。例如，如果一根木棍的长度是12英寸，而其他两根都只有1英寸，那么显然你无法让短的木棍在中间相接。对于任意三个长度，有一个简单的测试可以检验它们是否可能组成一个三角形：

如果其中有任意一个长度的值大于其他两个长度的和，则你不能组成三角形。否则可以。（如果两个长度的和等于第三个，则它们组成一个“退化”的三角。）

1. 编写一个函数`is_triangle`，接收三个整数参数，并根据这组长度的木棍是否能组成三角形来打印“Yes”或“No”。

2. 编写一个函数提示用户输入三根木棍的长度，将其转换为整数，并使用`is_triangle`检查这些长度的木棍是否可以组成三角形。

### 练习5-5

阅读下面的函数，并看看你能否弄清楚它在做什么。接着运行它（参看第4章中的示例）。

```
def draw(t, length, n):
    if n == 0:
        return
    angle = 50
    fd(t, length*n)
    lt(t, angle)
    draw(t, length, n-1)
    rt(t, 2*angle)
    draw(t, length, n-1)
    lt(t, angle)
    bk(t, length*n)
```

### 练习5-6

科赫曲线（Koch curve）是一个分形，它看起来像图 5-2所示。要绘制一个长度为 $x$ 的科赫曲线，你只需要做：

1. 绘制长度为 $x/3$ 的科赫曲线。
2. 向左转60度。
3. 绘制长度为 $x/3$ 的科赫曲线。
4. 向右转120度。
5. 绘制长度为 $x/3$ 的科赫曲线。
6. 向左转60度。
7. 绘制长度为 $x/3$ 的科赫曲线。

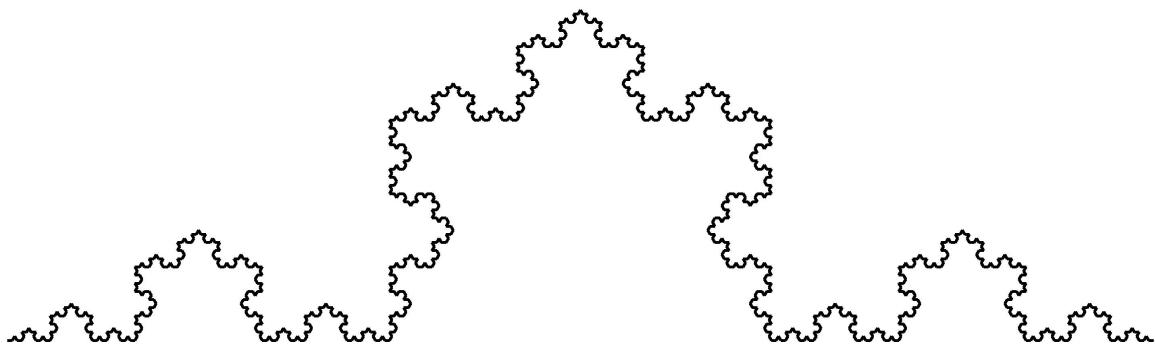


图5-2 一个科赫曲线

当 $x$ 比3小的时候例外：在那种情况下，你可以直接绘制一个长度为 $x$ 的直线。

1. 编写一个函数koch，接收一个Turtle以及一个长度作为形参，并使用Turtle绘制指定长度的科赫曲线。
2. 编写一个函数snowflake，绘制3条科赫曲线，组成一个雪花形状。解答：<http://thinkpython.com/code/koch.py>。
3. 科赫曲线可以用几种方法泛化。参看[http://en.wikipedia.org/wiki/Koch\\_snowflake](http://en.wikipedia.org/wiki/Koch_snowflake)中的例子，并实现你最喜欢的一个。

# 第6章 有返回函数

## 6.1 返回值

我们使用过的内置函数中，有一部分，例如数学函数，会返回结果。调用这些函数会产生一个值，我们将它赋值给一个变量或者用作表达式的组成部分。

```
e = math.exp(1.0)
height = radius * math.sin(radians)
```

至今为止我们写的函数都是无返回值函数。它们打印某些信息，或者移动乌龟，但它们的返回值是None。

本章中，我们会（终于）写一些有返回值函数。第一个例子是area，用于计算给定半径的圆的面积：

```
def area(radius):
    temp = math.pi * radius**2
    return temp
```

之前我们已经见过return语句，但在有返回值函数中，return语句包含了一个表达式。这个语句的意思是：“立即从这个函数中返回，并使用后面的表达式作为返回值。”表达式可以任意复杂，所以我们可以把这个函数写得更紧凑：

```
def area(radius):
    return math.pi * radius**2
```

另一方面，类似于temp的临时变量常常会让调试更容易。

有时候函数中针对不同的条件分支，各有各的返回语句会很有用处：

```
def absolute_value(x):
    if x < 0:
        return -x
    else:
        return x
```

因为return语句分别在不同的分支中，只有一个会被执行。

一旦return语句被执行，当前的函数就会终结，后面的语句不会执行。return语句之后的代码，或者在其他程序流程永远不可能达到的地方的代码，称为无效代码（dead code）。

在有返回函数中，保证每个可能执行路径上都会遇到return语句，是个很好的主意。例如：

```
def absolute_value(x):
    if x < 0:
        return -x
    if x > 0:
        return x
```

这个函数并不正确，因为如果x正好是0，则两个条件都不为true，则此时函数会没有遇到return语句就终结了。如果执行流程到了函数的结尾，返回值是None，它并不是0的绝对值。

```
>>> print absolute_value(0)
None
```

顺便说一下，Python内置提供了计算绝对值的函数abs。

### 练习6-1

写一个compare函数，如果x > y，返回1，如果x == y，返回0，如果x < y，返回-1。

## 6.2 增量开发

当你写大一些的函数时，可能会发现需要更多的时间来调试。为了对应不断增加的程序复杂度，你可能会想尝试一下称为增量开发的过程。增量开发的目标是通过每次只增加和测试一小部分代码，来避免长时间的调试过程。

例如，假设你想要查找两点之间的距离，给定坐标 $(x_1, y_1)$ 和 $(x_2, y_2)$ 。根据毕达哥拉斯定理，距离是：

$$\text{距离} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

第一步考虑Python中`distance`函数应该是什么样子的。换句话说，输入（形参）是什么？输出（返回值）是什么？

在这个例子中，输入是两个点，并可以用4个数字来表示。返回值是距离，它是一个浮点数。

现在就可以写出函数的轮廓了：

```
def distance(x1, y1, x2, y2):
    return 0.0
```

显然，现在这个版本计算的并不是距离；它总是返回0。但它是语法结构正确的，并且能运行，即意味着你可以在继续开发更复杂的功能之前对它进行初步的测试。

要测试这个新函数，使用样本参数调用它：

```
>>> distance(1, 2, 4, 6)
0.0
```

我选择这些值，因为这样两个点之间，横向距离是3，纵向距离是4；也就是，结果是5（3-4-5直角三角形的斜边）。当测试一个函数时，事先知道正确的结果是很有用的。

到这个时候我们已经确认函数的语法形式是正确的，紧接着可以给函数体添加代码了。合理的下一个步骤是找到距离差 $x_2 - x_1$ 和 $y_2 - y_1$ 。下一版本的函数将这两个距离差保存到临时变量中并打印出来。

```
def distance(x1, y1, x2, y2):
    dx = x2 - x1
    dy = y2 - y1
    print 'dx is', dx
    print 'dy is', dy
    return 0.0
```

如果函数正确执行，应该会显示‘`dx is 3`’和‘`dy is 4`’。如果确实如此，我们就可以确认函数正确地获得了实参，并正确地执行了第一步计算。如果不是如此，则只有几行代码需要检查。

下一步我们计算`dx`和`dy`的平方和：

```
def distance(x1, y1, x2, y2):
    dx = x2 - x1
    dy = y2 - y1
    dsquared = dx**2 + dy**2
    print 'dsquared is: ', dsquared
    return 0.0
```

同样地，你可以在这里再运行一遍程序，并检查输出（应该是25）。最后，可以使用`math.sqrt`来计算并返回结果：

```
def distance(x1, y1, x2, y2):
    dx = x2 - x1
    dy = y2 - y1
    dsquared = dx**2 + dy**2
    result = math.sqrt(dsquared)
    return result
```

如果这个函数运行正确，那么你的任务就完成了。否则，你可能需要在`return`语句之前打印出`result`的值。

最终版本的函数运行时并不打印任何东西；它只会返回一个值。我们之前写的`print`语句在调试时很有用，但一旦你的函数编写正确，就应该删除掉它们。这种代码称为脚手架代码（scaffolding），因为它们在构建程序的过程中很有用，但并不是最终产品的一部分。

开始的时候，应当每次只添加一到两行代码。当你获得更多经验之后，就会发现自己可以编写和调试更多的代码了。不论如何，增量开发都能帮你节省大量的调试时间。

这个过程的关键点是：

1. 以一个可以正确运行的程序开始，每次只做小的增量修改。如果在任意时刻发现错误，你都应当知道错误在哪里；
2. 使用临时变量保存计算的中间结果，这样你可以显示和检查它们；
3. 一旦整个程序完成，你可能会想要删除掉某些脚手架代码或者把多个语句综合到一个复杂表达式中。但只在不会增加代码阅读的难度时才应该那么做。

#### 练习 6-2

使用增量开发来编写一个函数 `hypotenuse`，给定直角三角形的两个直角边的长度时，它返回斜边的长度。开发过程中，记录每一步的情况。

## 6.3 组合

如你所料，在一个函数中可以调用其他函数。这个能力被称为组合。

作为示例，我们会写一个函数，它接收两个点，圆心和圆周上的一点，并计算圆的面积。

假设圆心保存在变量`xc`和`yc`中，而圆周上的点保存在`xp`和`yp`上。第一步是算出圆的半径，也就是这两个点的距离。我们刚才已经写了一个函数，`distance`，正好有这个功能：

```
radius = distance(xc, yc, xp, yp)
```

第二步是使用上一步算出来的半径来计算圆的面积；我们刚才也写了这个函数：

```
result = area(radius)
```

将这两步封装成一个函数，我们得到：

```
def circle_area(xc, yc, xp, yp):
    radius = distance(xc, yc, xp, yp)
    result = area(radius)
    return result
```

临时变量`radius`和`result`在开发和调试时有用，可一旦程序已经可以正确运行，我们就可以使用函数组合来简化函数：

```
def circle_area(xc, yc, xp, yp):
    return area(distance(xc, yc, xp, yp))
```

## 6.4 布尔函数

函数可以返回布尔值，这样可以很方便地隐藏函数内复杂的检测。例如：

```
def is_divisible(x, y):
    if x % y == 0:
        return True
    else:
        return False
```

通常布尔函数的命名都类似于是/否的问句。`is_divisible`返回 `True` 或 `False`，表示 `x` 是否可以被 `y` 整除。

这里是一个例子：

```
>>> is_divisible(6, 4)
False
>>> is_divisible(6, 3)
True
```

`==` 操作符的结果是一个布尔值，所以我们可以把这个函数写得更加紧凑：

```
def is_divisible(x, y):
    return x % y == 0
```

布尔函数常常用在条件语句中：

```
if is_divisible(x, y):
    print 'x is divisible by y'
```

你可能会想这么写：

```
if is_divisible(x, y) == True:
    print 'x is divisible by y'
```

但这里多出来的比较是不必要的。

练习6-3

写一个函数 `is_between(x, y, z)`，当  $x \leq y \leq z$  时，返回 `True`，其他情况返回 `False`。

## 6.5 再谈递归

至今为止，我们只涉及Python的一个很小的子集，但你可能会有兴趣知道，这个子集已经是一个完备的编程语言，也就是说，任何可以计算的问题，都可以用这个子集语言来完成。任何已有的程序，都可以用你现在已经学会的语言特性重写出来（实际上，可能还需要一些命令来控制诸如键盘、鼠标、光盘之类的设备，但仅此而已）。

要证明这个论断，并不是简单的工作。这个证明最早是由第一代计算机科学家之一阿兰·图灵（Alan Turing）完成的（有人会争辩他其实是个数学家，但大部分早期的计算机科学家都是从数学家开始的）。因此，这被称为图灵论题（Turing Thesis）。若想了解关于图灵论题的更完整（更准确）的讨论，我推荐Michael Sipser的《计算理论导引》（*Introduction to the Theory of Computation*）一书。

为了初步了解如何使用我们现在学会的工具，可以考虑几个递归定义的数学函数。递归定义和循环定义有些相似，因为同样地，定义中都会包含要定义的事物本身。真正的循环定义往往没什么用：

Vorpal:

一个形容词，用来描述一个vorpal的事物。

如果你在词典中看到这样的定义，可能会感觉恼怒。另一方面，如果你查看阶乘函数（用!表示）的定义，可能会看到如下内容：

```
0! = 1  
n! = n(n - 1)!
```

这个定义说明0的阶乘是1，而任意其他值 $n$ 的阶乘是 $n-1$ 的阶乘乘以 $n$ 。

所以 $3!$ 是3乘以 $2!$ ，而 $2!$ 是2乘以 $1!$ ，而 $1!$ 是1乘以 $0!$ 。综合起来， $3!$ 等于3乘以2乘以1乘以1，即6。

如果能够使用递归定义来描述一个事物，那么常常也可以使用Python程序来计算它。第一步是决定使用什么形参。在这个例子里，很明显函数factorial需要一个整数形参：

```
def factorial(n):
```

如果实参正好是0，我们只需要直接返回1：

```
def factorial(n):
    if n == 0:
        return 1
```

否则，接下来是有意思的地方，我们需要递归调用函数来计算 $n-1$ 的阶乘，并乘以 $n$ ：

```
def factorial(n):
    if n == 0:
        return 1
    else:
        recurse = factorial(n-1)
        result = n * recurse
        return result
```

这个程序的运行流程和5.8节里的countdown函数类似。如果我们使用实参值3调用factorial：

因为3不是0，我们使用第二个分支，计算 $n-1$ 的阶乘……

因为2不是0，我们使用第二个分支，计算 $n-1$ 的阶乘……

因为1不是0，我们使用第二个分支，计算 $n-1$ 的阶乘……

因为0是0，我们使用第一个分支并返回1，不再需要进行任何递归调用了。

返回值（1）乘以 $n=1$ ，结果返回。

返回值（1）乘以 $n=2$ ，结果返回。

返回值（2）乘以 $n=3$ ，结果是6，而这个结果就是整个函数的返回值。

图6-1显示了这一系列函数调用的栈图。

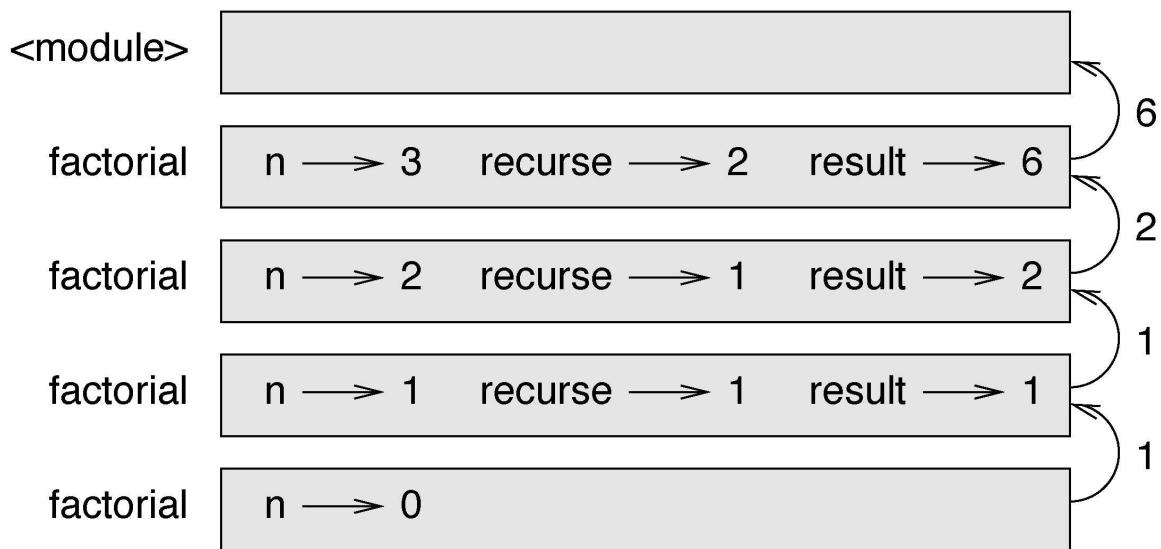


图6-1 栈图

结果值在图中显示为沿着栈向上传回。在每个帧中，返回值是result的值，即n和recurse的乘积。

最后一帧中，局部变量recurse和result不存在，因为新建它们的分支并没有运行。

## 6.6 坚持信念

跟踪程序执行的流程是阅读程序的一个办法，但那样很快就会陷入迷宫境况。另外有个办法，我称为“坚持信念”。当你遇到一个函数调用时，不去跟踪执行的流程，而假定函数是正确工作，能够返回正确的结果。

事实上，在使用内置函数时，你已经在这样尝试着坚持信念了。当调用`math.cos`或`math.exp`时，你并不去检查那些函数的内部实现。你只会假定它们是正确的，因为写这些内置函数的一定是很优秀的程序员。

当你调用自己写的函数时，这个道理也成立。比如，在6.4节中，我们写了`is_divisible`函数用来判断一个数是否可以被另一个数整除。一旦我们说服自己认定这个函数是正确的——通过检查代码和测试——就可以直接使用它，而不需要再细看内部实现了。

对递归函数来说，也是如此。当你调用递归函数时，不需要检查执行的流程，你应该假定递归调用是正确的（返回正确的结果），并问你自己，“假设我能够正确得到 $n-1$ 的阶乘，如何计算 $n$ 的阶乘？”在这个例子里，很明显你可以做到，直接乘以 $n$ 即可。

当然，在还没有完成函数的编写时，就假设它能正确工作，看起来有些奇怪，但那也正是为什么我称它为“坚持信念”的原因！

## 6.7 另一个示例

除阶乘factorial之外，最常见的递归数学定义是斐波那契数列(fibonacci)，其定义如下（参看[http://en.wikipedia.org/wiki/Fibonacci\\_number](http://en.wikipedia.org/wiki/Fibonacci_number)）：

$$\text{fibonacci}(0) = 0$$

$$\text{fibonacci}(1) = 1$$

$$\text{fibonacci}(n) = \text{fibonacci}(n - 1) + \text{fibonacci}(n - 2)$$

翻译成Python后，看起来是这样：

```
def fibonacci(n):
    if n == 0:
        return 0
    elif n == 1:
        return 1
    else:
        return fibonacci(n-1) + fibonacci(n-2)
```

如果你在这里试图跟踪执行的流程，即使是很小的参数 $n$ ，都会感觉头都要炸了。但因为坚持信念，如果你假定两个递归调用都正常工作，那么很明显，把它们加到一起必然得到正确的结果。

## 6.8 检查类型

如果我们调用factorial函数，并给定1.5作为实参，会发生什么呢？

```
>>> factorial(1.5)
RuntimeError: Maximum recursion depth exceeded
```

看起来像是无限递归。但怎么会这样？函数中有一个基准情形——当n == 0时。但如果n不是整数，我们就可能错过这个基准情形，而永远递归下去。

在第一个递归调用中，n是0.5。第二个，n是-0.5。从此以后，它会越来越小（更小的负数），但永远不会变成0。

我们有两个选择。可以尝试泛化函数factorial，使之能正确处理浮点数，或者我们也可以让factorial检查其实参的类型。第一个选择在数学上叫做伽玛函数（gamma function），它有些超出了本书的范围。所以我们选择第二个。

我们可以使用内置函数isinstance来检查实参的类型。与此同时，我们也可以确保实参是正数：

```
def factorial (n):
    if not isinstance(n, int):
        print 'Factorial is only defined for integers.'
        return None
    elif n < 0:
        print 'Factorial is not defined for negative integers.'
        return None
    elif n == 0:
        return 1
    else:
        return n * factorial(n-1)
```

第一个基准情形处理非整数；第二个捕获负数。这两种情况下，程序打印一个错误信息，并返回None，表示出现了问题：

```
>>> factorial('fred')
Factorial is only defined for integers.
None
>>> factorial(-2)
```

```
Factorial is not defined for negative integers.
```

```
None
```

如果我们通过了这两个测试，就能确保知道 $n$ 是正数或0，所以我们可以证明递归必然终结。

这个程序演示了一个模式，它有时被称为守卫（guardian）。前两个条件就像守卫一样，保护后面的代码，以免出现错误。守卫使得证明代码的正确性成为可能。

在11.3节中我们会看到一个更灵活的方案，用以打印错误信息：抛出一个异常。

## 6.9 调试

将一个大程序分解为小函数，自然而然地引入了调试的检查点。如果一个函数不能正常工作，可以考虑3种可能性。

- 函数获得的实参有问题，某个前置条件没有达到。
- 函数本身有问题，某个后置条件没有达到。
- 函数的返回值有问题，或者使用的方式不正确。

要排除第一种可能，你可以在函数开始的地方加上print语句，显示实参的值（以及它们的类型）。或者可以添加代码来显式地检查前置条件。

如果实参看起来没错，在每个return语句前添加print语句，显示返回值。如果有可能，手动检查返回值。考虑使用能更容易检验结果的实参来调用函数，就像6.2节中的那样。

如果函数看起来正常，检查调用它的代码，确保返回值被正确使用（或者确实被使用了！）。

在函数的开端和结尾处增加print语句，能帮助我们更清晰地了解函数的执行流程。例如，这里是一个添加了print语句的factorial函数：

```
def factorial(n):
    space = ' ' * (4 * n)
    print space, 'factorial', n
    if n == 0:
        print space, 'returning 1'
        return 1
    else:
        recurse = factorial(n-1)
        result = n * recurse
        print space, 'returning', result
    return result
```

space是一个字符串，包含多个空格，用来控制输出内容的缩进。下面是调用factorial(5)的结果：

```
factorial 5
factorial 4
factorial 3
factorial 2
```

```
    factorial 1
factorial 0
returning 1
    returning 1
        returning 2
            returning 6
                returning 24
                    returning 120
```

如果你对函数调用的流程有困惑，这种输出可以帮助你。开发有效的脚手架代码需要花费时间，但一点点脚手架可以节省大量的调试。

## 6.10 术语表

临时变量（temporary variable）：在复杂计算中用于保存中间计算值的变量。

无效代码（dead code）：程序中的一些代码，永远不可能执行到。常常是写在return语句之后的代码。

None：函数返回的一种特殊值，当没有返回语句，或者返回语句不带参数时，返回这个值。

增量开发（incremental development）：一个程序开发计划，通过每次只增加少量代码，并加以测试的步骤，来减少调试。

脚手架代码（scaffolding）：在开发过程中使用的，但在最终版本中不需要的代码。

守卫（guardian）：一个编程模式。使用条件语句来检查并处理可能产生错误的情形。

## 6.11 练习

### 练习6-4

为下面的程序绘制一个栈图。程序的输出是什么？

解答：[http://thinkpython.com/code/stack\\_diagram.py](http://thinkpython.com/code/stack_diagram.py)。

```
def b(z):
    prod = a(z, z)
    print z, prod
    return prod

def a(x, y):
    x = x + 1
    return x * y

def c(x, y, z):
    total = x + y + z
    square = b(total)**2
    return square

x = 1
y = x + 1
print c(x, y+3, x+y)
```

### 练习 6-5

Ackermann函数， $A(m, n)$ ，定义如下：

$$A(m, n) = \begin{cases} n + 1 & \text{if } m = 0 \\ A(m - 1, 1) & \text{if } m > 0 \text{ and } n = 0 \\ A(m - 1, A(m, n - 1)) & \text{if } m > 0 \text{ and } n > 0 \end{cases}$$

参考[http://en.wikipedia.org/wiki/Ackermann\\_function](http://en.wikipedia.org/wiki/Ackermann_function)。编写一个函数ack，计算Ackermann函数的值。使用你的函数求 $\text{ack}(3, 4)$ 的值，它应当是125。对于很大的数字m和n，会发生什么？

解答：<http://thinkpython.com/code/ackermann.py>。

### 练习6-6

回文是一个正向和逆向拼写都相同的单词，比如“noon”和“redivider”。递归地说，如果一个单词第一个和最后一个字母相同，并且中间是一个回文，则该单词是回文。

下面的函数接收一个字符串形参并返回第一个、最后一个以及中间的字母：

```
def first(word):
    return word[0]

def last(word):
    return word[-1]

def middle(word):
    return word[1:-1]
```

在第8章中查看它们是如何使用的。

1. 将这些函数保存到一个文件palindrome.py中并测试它们。如果你使用一个包含两个字母的字符串调用middle，会发生什么？使用一个字母呢？空字符串呢？空字符串写做' '并且不包含任意字母。
2. 编写一个函数is\_palindrome，接收一个字符串形参，并当它是回文的时候返回True，否则返回False。记着你可以使用内置函数len来检测字符串的长度。

解答：[http://thinkpython.com/code/palindrome\\_soln.py](http://thinkpython.com/code/palindrome_soln.py)。

#### 练习6-7

我们说一个数 $a$ 是 $b$ 的乘方，如果 $a$ 可以被 $b$ 整除，并且 $a/b$ 也是 $b$ 的乘方。编写一个函数is\_power接收形参 $a$ 和 $b$ ，当 $a$ 是 $b$ 的乘方时返回True。注意：你需要考虑基准情形。

#### 练习6-8

$a$ 和 $b$ 的最大公约数（GCD）是它们两个都能整除的最大的数。

寻找两个数的最大公约数的方法之一是使用欧几里得算法。这个算法基于如下观察：如果 $r$ 是 $a$ 除以 $b$ 的余数，则 $\gcd(a, b) = \gcd(b, r)$ 。作为基准情形，我们可以使用 $\gcd(a, 0) = a$ 。

编写一个函数gcd，接收形参a和b，并返回它们的最大公约数。如果你需要帮助，参看  
[http://en.wikipedia.org/wiki/Euclidean\\_algorithm](http://en.wikipedia.org/wiki/Euclidean_algorithm)。

鸣谢：这个练习是基于Abelson和Sussman的《计算机程序的结构和解释》（*Structure and Interpretation of Computer Programs*）一书。

# 第7章 迭代

## 7.1 多重赋值

你应当已经发现，对一个变量进行多次赋值是合法的。新的赋值语句使现有的变量引用一个新值（并不再引用老值）。

```
bruce = 5  
print bruce,  
bruce = 7  
print bruce
```

这段程序的输出是 5 7，因为第一次bruce打印时，它的值是5，而第二次时它的值是7。第一个print语句结尾的逗号用来避免换行，所以输出的结果在同一行上。

图7-1显示了在状态图中，多重赋值是什么样子的。

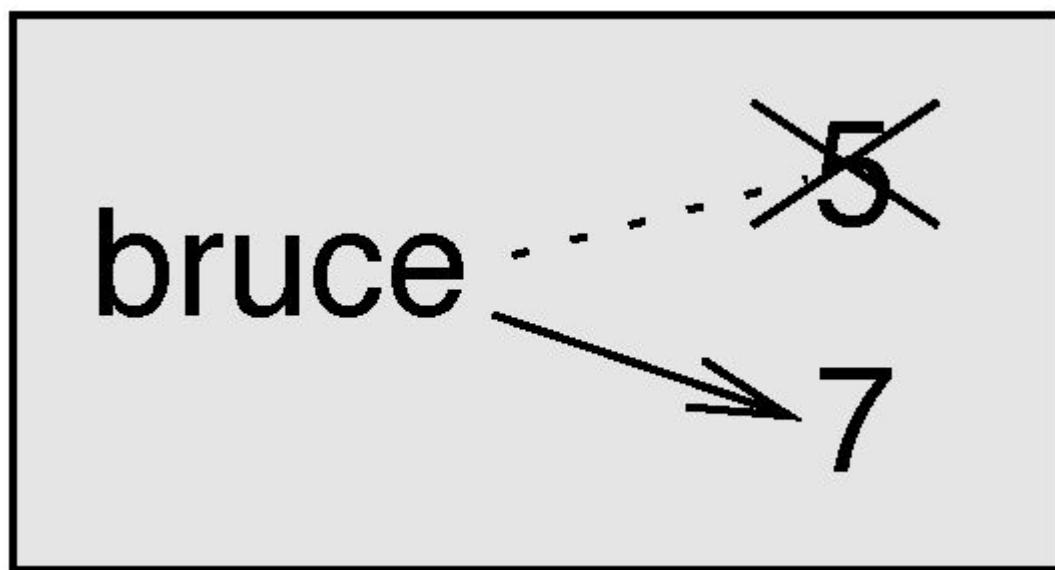


图7-1 状态图

因为有多重赋值，所以区分赋值操作符和相等操作符就非常重要。因为Python使用等号 (=) 来表示赋值，故很容易将a = b这样的

赋值语句错误理解为相等判断。它不是！

首先，相等判断是个对称的关系，而赋值并不是。比如，在数学中，如果 $a=7$ 那么 $7=a$ 。但是在Python中，语句`a = 7`是合法的，但`7 = a`则是非法的。

另外，在数学中，一个相等判断语句总是非真即假。如果现在 $a=b$ ，那么 $a$ 总会等于 $b$ 。在Python中，赋值语句会让两个变量变得相等，但它们并不会总保持那个状态：

```
a = 5
b = a      # a和b现在相等了
a = 3      # a和b不再相等了
```

第三行修改`a`的值，但是并不会修改`b`的值，所以它们不再相等。

虽然多重赋值常常很有用处，你应该谨慎使用它。如果变量的值经常变化，会导致程序难以阅读和调试。

## 7.2 更新变量

多重赋值的最常见形式是一个更新操作，即变量的新值依赖于旧值。

```
x = x+1
```

这个语句的意思是“获取x的当前值，加一，再更新x为此新值”。

如果你尝试更新一个并不存在的变量，则会得到错误，因为Python在赋值给x之前会先计算等号右边的部分：

```
>>> x = x+1
NameError: name 'x' is not defined
```

在更新变量之前，必须先对它进行初始化。通常通过一个简单赋值操作来进行初始化：

```
>>> x = 0
>>> x = x+1
```

通过加1来更新一个变量，称为增量（increment）；减1的操作称为减量（decrement）。

## 7.3 while语句

计算机常被用来自动化重复处理某些任务。重复相同或相似的任务，而不犯错误，这是电脑所擅长于人之处。

我们之前已经看到两个程序，`countdown`和`print_n`，它们使用递归来进行重复操作。这种重复被称为迭代（iteration）。因为迭代如此常见，Python提供了几个语言特性来支持它。其中一个是我们在4.2节中见过的`for`循环语句。后面我们会回到那个话题上。

另一个是`while`语句。下面是`countdown`函数的使用`while`语句实现的版本：

```
def countdown(n):
    while n > 0:
        print n
        n = n-1
    print 'Blastoff!'
```

你基本上可以按照英语来理解`while`语句。它的意思是，“每当n还大于0时，显示n的值，并将n减1。当n变成0的时候，显示单词`Blastoff!`”。

用更正式的说法，下面是`while`语句执行的流程：

1. 计算条件，得到`True`或`False`。
2. 如果条件为假，退出`while`语句，并继续执行后面的语句。
3. 如果条件为真，则执行`while`语句的语句体，并返回第1步。

这种类型的流程称为循环（loop），因为第3步又循环返回到最顶端的第1步了。

循环的语句体里面应当修改一个或多个变量的值，以致循环的条件最终能变成假，而退出循环。否则这个循环会永远重复下去，这样

的情况叫做无限循环。计算机科学家在读到洗发液的说明“涂抹、冲洗、重复”时，总会感到有趣，因为这是一个无限循环。

在countdown这个例子里，我们可以证明循环必然终结，因为我们知道n是有限的，并且可以看到每次循环，n的值都会减小，因此最终n会变成0。其他的情况下，就并不一定那么容易判断了：

```
def sequence(n):
    while n != 1:
        print n,
        if n%2 == 0:          # n是偶数
            n = n/2
        else:                 # n是奇数
            n = n*3+1
```

这个循环的条件是 $n \neq 1$ ，所以只要n还没有变成1而导致条件变假，循环就会一直进行下去。

每一个循环中，程序输出n的值，并检查它是偶数还是奇数。如果是偶数，n会除以2。如果是奇数，n会被替换为 $n*3+1$ 。比如，如果传入sequence函数的参数是3，则结果序列是：3，10，5，16，8，4，2，1。

因为n有时候增加，有时候减少，没有办法找到明显的证据确定n一定会最终变成1，或者说程序会终止。对于某些特定的n值，我们可以证明最终会终止。例如，如果开始的参数值是2的幂方，则每次循环都会进入偶数分支，直到变成1。前面的例子中有一部分就是这样的序列，以16开始。

但困难的问题是，我们是否能够证明这个程序对所有的正值n都可以最终终止。至今为止，还没有人对这个问题给出证明或证伪！（参考[http://en.wikipedia.org/wiki/Collatz\\_conjecture](http://en.wikipedia.org/wiki/Collatz_conjecture)）。

### 练习7-1

重写5.8节中的print\_n函数，使用循环而非递归来实现。

## 7.4 break语句

有时候只有在循环语句体的执行途中才能知道是不是到了退出循环的时机。这时候可以使用break语句来跳出循环。

例如，假设你想要获得用户输入，直到他们输入done。可以这么写：

```
while True:
    line = raw_input('> ')
    if line == 'done':
        break
    print line
print 'Done!'
```

循环的条件是True，总是为真，所以循环会一直运行，直到遇到break语句。

每次循环之内，都会先用一个‘>’来提示用户输入。如果用户输入done，break语句会退出循环。否则程序会显示出用户输入的内容，并重新回到循环的顶端。这里是一个运行的实例：

```
> not done
not done
> done
Done!
```

这种写while循环的方式很常见，因为可以把判断循环条件的逻辑放在循环中的任何地方（而不只是在顶端），并且可以以肯定的语气来表示终结条件（“当这样发生时停止循环”），而不是否定的语气（“继续执行，直到那个条件发生”）。

## 7.5 平方根

程序中常常使用循环来进行数值计算，以一个近似值开始，并迭代地优化计算结果。

例如，计算平方根的方法之一是牛顿方法。假设你想要知道 $a$ 的平方根。如果你以任意一个估计值 $x$ 开始，可以使用如下的方程获得一个更好的估计值。

$$y = \frac{x + a/x}{2}$$

例如，如果 $a$ 是4而 $x$ 是3：

```
>>> a = 4.0
>>> x = 3.0
>>> y = (x + a/x) / 2
>>> print y
2.16666666667
```

这个值更接近正确的答案 ( $\text{sqrt}(4) == 2$ )。如果我们使用新的估计值重复这个过程，会得到更近似的结果：

```
>>> x = y
>>> y = (x + a/x) / 2
>>> print y
2.00641025641
```

再经过几次重复更新，估计值会几乎完全准确：

```
>>> x = y
>>> y = (x + a/x) / 2
>>> print y
2.00001024003
>>> x = y
>>> y = (x + a/x) / 2
>>> print y
2.00000000003
```

通常来说，我们并不能提前知道需要多少步才能得到正确的答案，但是当估计值不再变化时，我们就知道达到目的了。

```
>>> x = y
>>> y = (x + a/x) / 2
>>> print y
2.0
>>> x = y
>>> y = (x + a/x) / 2
>>> print y
2.0
```

当 $y == x$ 时，可以终止。下面是一个以估计值 $x$ 开始，并不断迭代优化直到它不再变化的循环：

```
while True:
    print x
    y = (x + a/x) / 2
    if y == x:
        break
    x = y
```

对于大多数 $a$ 值来说，这样效果很好，但通常来说，测试`float`的相等是危险的。浮点数值只是近似正确：大部分有理数，比如 $1/3$ ，以及无理数，如 $\sqrt{2}$ ，都不能用`float`精确表示。

比起判断 $x$ 和 $y$ 是否精确相等，更安全的方式是利用内置函数`abs`来计算它们之间差值的绝对值，或者说量级：

```
if abs(y-x) < epsilon:
    break
```

这里`epsilon`的值是`0.0000001`，用来决定近似度是足够的。

## 练习7-2

将这个循环封装到一个函数`square_root`中，接收一个形参 $a$ 。选择一个合理的值 $x$ ，并返回 $a$ 的平方根的估计值。

## 7.6 算法

牛顿方法是算法的一个例子：它是解决一类问题的机械化过程（在这个例子里，问题是计算平方根）。

定义算法并不容易。从一个算不上算法的东西开始可能更简单。当你学习个位数相乘时，可能背诵过乘法表。实际上，你记住了100个特别的解答。这种知识不算是算法。

但如果你比较“懒”，可能会使用一些小技巧来偷懒。比如，要计算 $n$ 和9的乘积，你可以写下 $n-1$ 作为十位数， $10-n$ 作为个位数。这个小技巧是计算任意个位数和9的乘积的通用方案。这算是一个算法！

相似地，你学习过的进位加法、借位减法以及长除法都是算法。算法的特点之一是它们不需要任何聪明才智就能执行。它们是一个机械化的过程，每一步都依照一组简单的规则接着上一步进行。

在我看来，人类花费如此多的时间在学校学习如何执行算法，有些尴尬。执行它们，非常直白，不需要任何智力。

另一方面，设计算法的过程则充满趣味和智力挑战。它是我们称为程序设计的核心成分。

一些人们自然而然、毫无困难或者下意识所做的事情，用算法表达却最为困难。理解自然语言是一个好例子。我们都能理解自然语言，但是至今为止还没有人能解释我们是怎么做到的，至少没办法用算法解释。

## 7.7 调试

当你开始编写更大的程序时，常常会发现自己花费更多的时间用于调试。更多的代码意味着更多的出错机会，以及更多可能隐藏着bug的地方。

削减调试时间的方法之一是“二分调试”（debugging by bisection）。例如，如果你的程序有100行代码，如果每次检查一行，需要100步。

相反地，可以尝试把问题分成两半。找到程序的中点，或者接近那里的地方，找一个可以检验的中间结果。添加一个print语句（或者其他可以有检查效果的代码）并运行程序。

如果中点检验的结果是错误的，说明错误必然出现在程序的前半部分。如果是正确的，那错误则在程序的后半部分。

每进行一次这样的检查，就减少了一半需要检查的代码。经过6步之后（显然少于100步），就能够减少到一至两行代码，至少理论上如此。

实践中，常常很难确定“程序的中点”在哪里，并且并不总是能够检验它。通过数代码行数来确定中点显然没有意义。相反地，应当思考程序中哪些地方可能出错，哪些地方容易加上一个检查。然后选择一个你认为在其前后发生错误几率差不多的点进行检查。

## 7.8 术语表

多重赋值 (multiple assignment)：在程序的执行过程中对同一个变量进行多次赋值。

更新 (update)：一种赋值操作，新值依赖于变量的旧值。

初始化 (initialization)：一种赋值操作，给变量一个初始的值，以后可以进行更新。

增量 (increment)：一种更新操作，增加变量的值（常常是加1）。

减量 (decrement)：一种更新操作，减少变量的值。

迭代 (iteration)：使用递归函数调用或者循环来重复执行一组语句。

无限循环 (infinite loop)：一个终止条件永远无法满足的循环。

## 7.9 练习

### 练习7-3

要测试本章中的平方根算法，你可以将它和math.sqrt比较。编写一个函数test\_square\_root，打印下面这样的表格：

1.0	1.0	1.0	0.0
2.0	1.41421356237	1.41421356237	2.22044604925e-16
3.0	1.73205080757	1.73205080757	0.0
4.0	2.0	2.0	0.0
5.0	2.2360679775	2.2360679775	0.0
6.0	2.44948974278	2.44948974278	0.0
7.0	2.64575131106	2.64575131106	0.0
8.0	2.82842712475	2.82842712475	4.4408920985e-16
9.0	3.0	3.0	0.0

第一列是一个数， $a$ ；第二列是数 $a$ 的平方根，使用本章中7.5节的函数计算；第三列是使用math.sqrt计算出的平方根；第四列是两种计算结果的差值的绝对值。

### 练习7-4

内置函数eval接收一个字符串并使用Python解释器对它进行求值。例如：

```
>>> eval('1 + 2 * 3')
7
>>> import math
>>> eval('math.sqrt(5)')
2.2360679774997898
>>> eval('type(math.pi)')
<type 'float'>
```

编写一个函数eval\_loop，迭代地提示用户，接收他们的输入并使用eval求值，并打印出结果。

它应当一直继续，直到用户输入'done'，并返回最后一个求值的表达式的结果。

## 练习7-5

数学家拉马努金（Srinivasa Ramanujan）找到了一个无限序列，可以用来生成 $\pi$ 的数值近似值：

$$\frac{1}{\pi} = \frac{2\sqrt{2}}{9801} \sum_{k=0}^{\infty} \frac{(4k)!(1103 + 26390k)}{(k!)^4 396^{4k}}$$

编写一个函数estimate\_pi，使用这个公式计算并返回 $\pi$ 的近似估计。它应当使用一个while循环来计算求和的每一项，直到最后一项的值小于 $1e-15$ （这是Python对 $10^{-15}$ 的标记法）。你可以通过和math.pi比较来检查计算的结果。

解答：<http://thinkpython.com/code/pi.py>。

# 第8章 字符串

## 8.1 字符串是一个序列

字符串是一个字符的序列（sequence）。可以使用方括号操作符来访问字符串中单独的字符：

```
>>> fruit = 'banana'  
>>> letter = fruit[1]
```

第二个语句选择`fruit`中的第1个字符，并将它赋值给`letter`变量。

方括号中的表达式称为下标（index）。下标表示你想要序列中的哪一个字符（所以用`index`这个名称）。

但你可能发现得到的和预料不一样：

```
>>> print letter  
a
```

对大多数人来说，'banana'的第一个字母是b，而不是a。但对计算机科学家来说，下标表示的是离字符串开头的偏移量，而第一个字母的偏移量是0。

```
>>> letter = fruit[0]  
>>> print letter  
b
```

所以b是'banana'的第0个字母，a是第1个，n是第2个。

任何表达式，包括变量和操作符，都可以作为下标。但下标的值必须是整数。否则你会得到：

```
>>> letter = fruit[1.5]  
TypeError: string indices must be integers
```

## 8.2 len

`len`是一个内置函数，返回字符串中字符的个数：

```
>>> fruit = 'banana'  
>>> len(fruit)  
6
```

要获得字符串的最后一个字母，你可能会想这么写：

```
>>> length = len(fruit)  
>>> last = fruit[length]  
IndexError: string index out of range
```

`IndexError`出现的原因是`'banana'`中没有下标为6的字母。因为我们是从0开始计算的，6个字母的下标是0到5。要获得最后一个字符，需要从`length`里减1：

```
>>> last = fruit[length-1]  
>>> print last  
a
```

或者，你可以使用负数下标。负数下标从字符串结尾处倒着数。表达式`fruit[-1]`返回最后一个字母，表达式`fruit[-2]`返回倒数第二个字母，依此类推。

## 8.3 使用for循环进行遍历

有很多计算都涉及对字符串每次处理一个字符的操作。它们常常从开头起，每次选择一个字符，对它做一些处理，再继续，直到结束。这种处理的模式，我们称为遍历（traversal）。编写遍历逻辑的方法之一是使用while循环：

```
index = 0
while index < len(fruit):
    letter = fruit[index]
    print letter
    index = index + 1
```

这个循环遍历字符串，并将每个字符显示在单独的一行上。循环的结束条件是`index < len(fruit)`，所以当`index`等于字符串的长度时，条件为假，循环体不被执行。最后访问的字符下标为`len(fruit) - 1`，正好是字符串最后一个字符。

### 练习8-1

写一个函数，接收一个字符串作为形参，并倒序显示它的字母，每个字母单独一行。

写遍历逻辑的另一个方式是使用for循环：

```
for char in fruit:
    print char
```

每次迭代之中，字符串中的下一个字符会被赋值给变量`char`。循环会继续直到没有剩余的字符为止。

下面的示例展示了如何利用字符串拼接（字符串加法）和一个for循环来生成字母序列（也就是，按字母顺序排序的序列）。在Robert McCloskey的书《为小鸭让路》（*Make Way for Ducklings*）中，小鸭们的名字是Jack、Kack、Lack、Mack、Nack、Ouack、Pack及Quack。下面的循环按顺序输出这些名字：

```
prefixes = 'JKLMNOPQ'
suffix = 'ack'
```

```
for letter in prefixes:  
    print letter + suffix
```

输出是：

```
Jack  
Kack  
Lack  
Mack  
Nack  
Oack  
Pack  
Qack
```

当然那并不完全正确，因为“Ouack”和“Quack”拼写错了。

练习8-2

修改程序解决这个问题。

## 8.4 字符串切片

字符串中的一段称为一个切片（slice）。选择一个切片和选择一个字符类似：

```
>>> s = 'Monty Python'  
>>> print s[0:5]  
Monty  
>>> print s[6:12]  
Python
```

操作符  $[n:m]$  返回字符串从第  $n$  个字符到第  $m$  个字符的部分，包含第  $n$  个字符，但不包含第  $m$  个字符。这个行为有些违反直觉，但如果想象下标是指向字符之间的位置，可以帮助我们理解它，如图8-1所示。

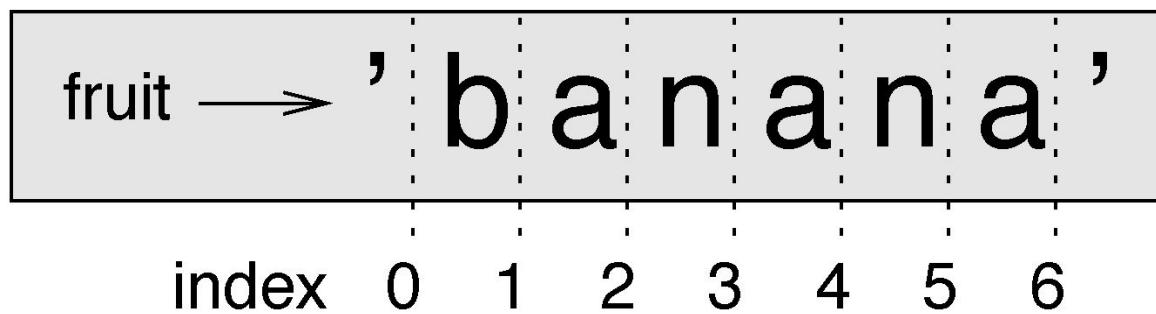


图8-1 切片的下标

如果省略掉第一个下标（冒号之前的那个），切片会从字符串开头开始。如果省略掉第二个下标，切片会继续到字符串的结尾。

```
>>> fruit = 'banana'  
>>> fruit[:3]  
'ban'  
>>> fruit[3:]  
'ana'
```

如果第一个下标大于或等于第二个下标，结果是空字符串，用两个引号表示：

```
>>> fruit = 'banana'  
>>> fruit[3:3]  
''
```

空字符串不包含任何字符，长度为0，但除此之外，它和其他字符串一样。

### 练习8-3

给定`fruit`是一个字符串，`fruit[:]`表示什么？

## 8.5 字符串是不可变的

想要修改字符串的某个字符，你可能会想直接在赋值左侧使用`[]`操作符。例如：

```
>>> greeting = 'Hello, world!'
>>> greeting[0] = 'J'
TypeError: object does not support item assignment
```

这个例子里的“对象”（object）是字符串，而“项”（item）是指你想要赋值的那个字符。就现在来说，一个对象和值是差不多的东西，但我们在后面细谈它。而项则是值序列中的一个。

这个错误产生的原因是因为字符串是不可变（immutable）的，也就是说，你不能修改一个已经存在的字符串。你能做的最多是新建一个字符串，它和原来的字符串稍有不同：

```
>>> greeting = 'Hello, world!'
>>> new_greeting = 'J' + greeting[1:]
>>> print new_greeting
Jello, world!
```

这个例子使用新的首字符和greeting的一个切片拼接起来。它对原来的字符串没有影响。

## 8.6 搜索

下面的这段函数是做什么的？

```
def find(word, letter):
    index = 0
    while index < len(word):
        if word[index] == letter:
            return index
        index = index + 1
    return -1
```

从某种意义上说，`find`是`[]`操作符的反面。和`[]`操作符通过一个下标查找对应的字符不同，它根据一个字符查找其出现在字符串中的下标。如果没有找到字符，函数返回`-1`。

这是我们第一次在循环内部看到`return`语句。如果`word[index] == letter`，函数直接跳出循环并立即返回。

如果字符没有出现在字符串中，程序正常退出循环，并返回`-1`。

这种计算的模式——遍历一个序列，并当找到我们寻找的目标时返回——称为**搜索**。

练习8-4

修改`find`函数，让它接收第3个形参，表示从`word`的那个下标开始搜索。

## 8.7 循环和计数

下面的代码计算字母a在字符串中出现的次数：

```
word = 'banana'  
count = 0  
for letter in word:  
    if letter == 'a':  
        count = count + 1  
print count
```

这个程序展示了另一种计算模式，称为计数器。变量count初始化为0，接着每次找到一个a时计数器加1。当循环结束时，count保存着结果——a出现的总次数。

练习8-5

将这段代码封装成函数count，并泛化它以接收字符串和要计数的字母作为形参。

练习8-6

重写count函数，不直接遍历字符串，而是使用前面一节中的3形参版本的find函数。

## 8.8 字符串方法

方法和函数很相似——它接收形参并返回值——但语法有所不同。例如，方法upper接收一个字符串，并返回一个全部字母都是大写的字符串。

和函数的语法upper(word)不同，它使用方法的语法  
word.upper()。

```
>>> word = 'banana'  
>>> new_word = word.upper()  
>>> print new_word  
BANANA
```

这种句点表示法指定了方法的名称，以及方法应用到的字符串的名称word。空的括号表示这个方法没有任何参数。

方法的调用称为invocation；在这个例子里，我们说我们在word字符串上调用方法upper。

实际上，字符串本来就有一个方法find，和我们之前写的find函数非常相似：

```
>>> word = 'banana'  
>>> index = word.find('a')  
>>> print index  
1
```

在这个例子中，我们在word上调用find方法，并传入要查找的字母作为实参。

实际上，find方法比我们的函数更通用；它可以用来查找子字符串，而不仅仅是字符：

```
>>> word.find('na')  
2
```

它还可以接收第二个实参，表示从哪一个下标开始查找：

```
>>> word.find('na', 3)
4
```

以及第三个实参，表示查找到哪个下标就结束：

```
>>> name = 'bob'
>>> name.find('b', 1, 2)
-1
```

这个搜索失败，因为b并没有在字符串的下标1到2之间（不包括2）出现。

#### 练习8-7

字符串有一个方法叫做count，和我们之前练习中的方法类似。阅读这个方法的文档，并写一个程序调用它来计算'banana'中a出现的次数。

#### 练习8-8

在<http://docs.python.org/lib/string-methods.html>阅读字符串方法的文档。你可能会想实验一下其中的一些方法，以确保理解它们如何工作。strip和replace特别有用。

文档中使用了一种可能会引起困惑的语法。例如，`find(sub[, start[, end]])`中的中括号表示可选的参数。所以`sub`是必需的，但是`start`是可选的，并且如果你使用了`start`，则`end`是可选的。

## 8.9 操作符in

in是一个布尔操作符，操作于两个字符串上，如果第一个是第二个的子串，则返回True，否则返回False：

```
>>> 'a' in 'banana'  
True  
>>> 'seed' in 'banana'  
False
```

例如，下面的函数打印出word1中出现且出现在word2中的所有字母：

```
def in_both(word1, word2):  
    for letter in word1:  
        if letter in word2:  
            print letter
```

精心选择变量名称后，Python有时会读起来很像英语。你可以这样阅读这个循环：“for (each) letter in (the first) word, if (the) letter (appears) in (the second) word, print (the) letter”。

下面是用这个函数比较单词apples和oranges的结果：

```
>>> in_both('apples', 'oranges')  
a  
e  
s
```

## 8.10 字符串比较

关系操作符也可以用在字符串上。检查两个字符串是否相等：

```
if word == 'banana':  
    print 'All right, bananas.'
```

其他的关系操作符在将单词按照字母顺序比较时有用：

```
if word < 'banana':  
    print 'Your word, ' + word + ', comes before banana.'  
elif word > 'banana':  
    print 'Your word, ' + word + ', comes after banana.'  
else:  
    print 'All right, bananas.'
```

Python处理大小写字母时和人不一样。所有的大写字母都在小写字母之前。所以：

```
Your word, Pineapple, comes before banana.
```

处理这个问题的常用办法是先将字符串都转换为标准的形式，如都转换成全小写字母形式，再进行比较。如果你遇到一个武装着Pineapple的敌人需要保护自己时，请记住这个办法。

## 8.11 调试

当使用下标来遍历序列中的值时，要正确实现遍历的开端和结尾并不容易。

下面是一个函数，能够比较两个单词，如果它们互为倒序，则返回True，但这个函数包含了两个错误：

```
def is_reverse(word1, word2):
    if len(word1) != len(word2):
        return False

    i = 0
    j = len(word2)

    while j > 0:
        if word1[i] != word2[j]:
            return False
        i = i+1
        j = j-1

    return True
```

第一个if语句检查两个单词是否长度相同。如果不同，我们就立即返回False，接着在后面整个函数中，都可以认为两个单词是相同长度的。这是6.8节中讲到的守卫模式的一个实例。

i和j是下标：i用于正向遍历word1，而j用于反向遍历word2。如果我们找到两个不匹配的字母，则可以立即返回False。如果完成整个循环后所有的字母仍然都相等，则返回True。

如果使用单词“pots”和“stop”来测试这个函数，我们会预期返回值是True，但实际上会得到一个IndexError：

```
>>> is_reverse('pots', 'stop')
...
File "reverse.py", line 15, in is_reverse
    if word1[i] != word2[j]:
IndexError: string index out of range
```

为了调试这类错误，第一步可以在发生错误的那行代码之前打印出索引的值。

```
while j > 0:  
    print i, j      # 在这里打印  
  
    if word1[i] != word2[j]:  
        return False  
    i = i+1  
    j = j-1
```

这样再一次运行程序时，能获得更多的信息：

```
>>> is_reverse('pots', 'stop')  
0 4  
...  
IndexError: string index out of range
```

第一次迭代时，`j`的值是4，超出了'pots'的范围。最后一个字符的下标是3，所以`j`的初始值应该是`len(word2)-1`。

如果修改这个错误并重新运行程序，会得到：

```
>>> is_reverse('pots', 'stop')  
0 3  
1 2  
2 1  
True
```

这回我们得到了正确的结果，但看起来循环只运行了3次，有些可疑。为了对具体发生了什么有更清晰的印象，可以画一个状态图。第一个迭代中，`is_reverse`的帧显示在图8-2中。

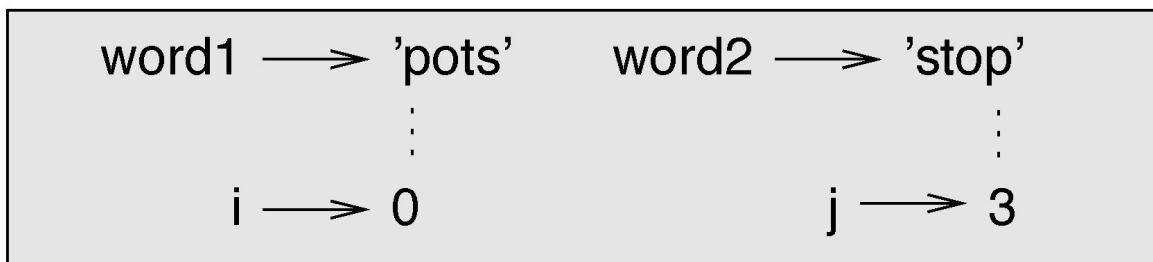


图8-2 状态图

我特意安排了帧中变量的位置，并使用虚线来显示`i`和`j`指向`word1`和`word2`中的字符。

练习8-9

从这个图开始，在纸上执行程序，每个迭代修改 $i$ 和 $j$ 的值。找到并修复这个函数的第二个错误。

## 8.12 术语表

**对象 (object)**：变量可以引用的一种事物。就现在来说，你可以把“对象”当作“值”来使用。

**序列 (sequence)**：一个有序的集合；也就是说，一组值，其中每个使用一个下标来定位。

**项 (item)**：序列中的一个值。

**下标 (index)**：用于在序列中选择元素的整数值。例如，可以用于在字符串中选取字符。

**切片 (slice)**：字符串的一部分，通过一个下标范围来定位。

**空字符串 (empty string)**：没有字符，长度为0的字符串，使用一对引号来表示。

**不可变 (immutable)**：序列的一种属性，表示它的元素是不可以改变的。

**遍历 (traverse)**：迭代访问序列中的每一个元素，并对每个元素进行相似的操作。

**搜索 (search)**：一种遍历的模式，当找到它想要的元素时停止。

**计数器 (counter)**：一种用来计数的变量，通常初始化为0，后来会递增。

**方法 (method)**：和对象关联的一个函数，使用句点表示法调用。

**方法调用 (invocation)**：调用一个方法的语句。

## 8.13 练习

### 练习8-10

字符串切片可以接受第三个下标用来指定“步长”，即相邻的字符之间的距离。步长为2，意思是切片每次取接下来第2个字符；步长3意思是每次取接下来第3个字符，等等。

```
>>> fruit = 'banana'  
>>> fruit[0:5:2]  
'bnn'
```

步长为-1表示切片按照相反的方向访问字符串，所以切片`[::-1]`会得到一个逆序的字符串。

使用这个特性来编写一个一行版本的`is_palindrome`函数（见练习6-6）。

### 练习8-11

下面的几个函数目的都是检查一个字符串是否包含小写字母，但至少有一个是错误的。对每个函数，描述一下这个函数到底做了什么（假设形参是一个字符串）。

```
def any_lowercase1(s):  
    for c in s:  
        if c.islower():  
            return True  
        else:  
            return False  
  
def any_lowercase2(s):  
    for c in s:  
        if 'c'.islower():  
            return 'True'  
        else:  
            return 'False'  
  
def any_lowercase3(s):  
    for c in s:  
        flag = c.islower()  
    return flag  
  
def any_lowercase4(s):  
    flag = False
```

```
for c in s:  
    flag = flag or c.islower()  
return flag  
  
def any_lowercase5(s):  
    for c in s:  
        if not c.islower():  
            return False  
    return True
```

### 练习8-12

ROT13是一个比较弱的加密形式，它涉及将单词中的每个字母“轮转”13个位置。轮转一个字母意思是在字母表中移动它，如果需要，从开头开始。所以‘A’移动3个位置是‘D’，而‘Z’移动一个位置是‘A’。

编写一个函数rotate\_word，接收一个字符串以及一个整数作为参数，并返回一个新字符串，其中的字母按照给定的整数值“轮转”位置。

例如，“cheer”轮转7位的结果是“jolly”，而“melon”轮转-10位结果是“cubed”。

你可以使用内置函数ord，它能够将一个字符转换为数值编码，以及函数chr，它将数值编码转换为字符。

因特网上有些可能侵犯人的笑话是用ROT13编码的。如果你不容易被冒犯，可以寻找一些并解码。解答：

<http://thinkpython.com/code/rotate.py>。

# 第9章 案例分析：文字游戏

## 9.1 读取单词列表

为本章的练习，我们需要准备一个英文单词列表。互联网上有很多可用的单词列表，但最适合我们的目标的单词列表，是由Grady Ward收集整理并作为Moby词典项目（参看[http://wikipedia.org/wiki/Moby\\_Project](http://wikipedia.org/wiki/Moby_Project)）的一部分贡献给公共域的。它包含113 809个正式的填字游戏用词，即那些认为可以用于纵横填字游戏和其他类型文字游戏的单词。在Moby集合中，文件名是113809of.fic；你可以从<http://thinkpython.com/code/words.txt>下载一个副本，但文件名是words.txt。

这个文件是纯文本，所以可以使用文本编辑器打开，也可以使用Python读入它。内置函数open接收文件名作为参数，并返回一个文件对象（file object），可以用来读取文件。

```
>>> fin = open('words.txt')
>>> print fin
<open file 'words.txt', mode 'r' at 0xb7f4b380>
```

fin是用来表示文件对象作为输入源时常用的名称。模式'r'表示文件是为了读取而打开的（相反地，'w'表示为了写入而打开）。

文件对象提供了几个方法用于读取内容，包括readline，它会从文件里读入字符，直到获得换行符为止，并将读入的结果作为一个字符串返回：

```
>>> fin.readline()
'aa\r\n'
```

在这个从words.txt读入的特定的列表中，第一个单词是"aa"，它是一种火山熔岩。序列\r\n表示两个空格字符，一个是回车，一个是换行，用于把这个单词和其他单词分隔开。

文件对象会记录它读到文件的那个位置，因此如果再次调用 `readline`，会得到下一个单词：

```
>>> fin.readline()
'aah\r\n'
```

下一个单词是”aah”，也是一个完全合法的单词，所以别用奇怪的眼光看着我。或者，如果是那几个空白字符在干扰你，可以使用字符串的方法`strip`去掉它们：

```
>>> line = fin.readline()
>>> word = line.strip()
>>> print word
aahed
```

你也可以在`for`循环中使用文件对象。下面的代码读入 `words.txt` 并每行打印出一个单词：

```
fin = open('words.txt')
for line in fin:
    word = line.strip()
    print word
```

### 练习9-1

编写一个程序，读入 `words.txt` 并且打印出那些长度超过20个字符的单词（不算空白字符）。

## 9.2 练习

在下一节里有这些练习的解答。在继续阅读解答之前，应当至少尝试一下每一个练习。

### 练习9-2

1939年，Ernest Vincent Wright出版了一本5万字的小说《Gadsby》，这本书里没有包含字母“e”。因为“e”是英语中最常见的字母，所以这并不是件容易的事。

实际上，不使用这最常见的字母的话，仅仅是构建一条单独的构思也是很难的事情。开始时会很慢而艰难，但保持谨慎和长时间的训练，你可以渐渐掌握方法。

好吧，我先停下来。<sup>①</sup>

写一个函数has\_no\_e，当给定的单词不包含字母“e”时，返回True。

修改前一节练习中的代码，打印出不含“e”的单词，并计算这种单词在整个单词表中的百分比。

### 练习9-3

编写一个函数avoids，接收一个单词，以及一个包含禁止字母的字符串，当单词不含任何禁止字母时，返回True。

修改你的程序，提示用户输入包含禁止字母的字符串，并打印出不包含任意禁止字母的单词的个数。能不能找到一组5个禁止字母的组合，它们排除的单词最少？

### 练习9-4

编写一个名为uses\_only的函数，接收一个单词以及字母组成的字符串，当单词只由这些字母组成时返回True。你可以造一个句子，其单词只由字母acefhlo组成吗？除了“Hoe alfalfa”之外呢？

### 练习9-5

编写一个名为uses\_all的函数，接收一个单词以及由需要的字母组成的字符串，当单词中所有需要的字母都出现了至少一次时返回True。有多少单词使用了所有的元音字母aeiou？而aeiouy呢？

### 练习9-6

编写一个名叫is\_abecedarian的函数，如果单词中的字母是按照字母表顺序排列的（两个重复字母也OK），则返回True。有多少这样的单词？

## 9.3 搜索

前面一节的所有练习都有一个共同点：它们可以使用我们在8.6节中介绍的搜索模式来解决。最简单的例子是：

```
def has_no_e(word):
    for letter in word:
        if letter == 'e':
            return False
    return True
```

for循环遍历单词word中的字符。如果我们找到字母“e”，可以立即返回False；否则只能继续下一个字母。如果正常退出了循环，则说明我们没有找到“e”，所以返回True。

avoids是has\_no\_e的更通用的版本，它们的结构相同：

```
def avoids(word, forbidden):
    for letter in word:
        if letter in forbidden:
            return False
    return True
```

一旦发现一个禁止的字母，可以立即返回False；如果运行到循环结束，则返回True。

uses\_only函数也类似，只是它条件判断的意思是相反的：

```
def uses_only(word, available):
    for letter in word:
        if letter not in available:
            return False
    return True
```

它接收的参数并不是一个禁止字母列表，而是一个可用字母列表available。如果我们发现单词中遇到了并不属于available的字母，则可以返回False。

`uses_all`函数也类似，但单词和字母列表的角色相反。

```
def uses_all(word, required):
    for letter in required:
        if letter not in word:
            return False
    return True
```

我们不再遍历单词`word`中的字母，而是循环遍历必需的单词列表`required`。如果单词列表中有任意字母没有出现在单词中，我们可以返回`False`。

如果你真的像计算机科学家那样思考的话，应该已经发现，`uses_all`实际上已经是解决的问题的一个特例，并且可以这么写：

```
def uses_all(word, required):
    return uses_only(required, word)
```

这是被称为问题识别（problem recognition）的程序开发方法的一个例子。意思是识别出的当前问题是一个已经解决的问题的特例，从而可以直接利用已经开发好的解决方案。

## 9.4 使用下标循环

在前面一节的例子中，我使用for循环进行遍历，因为只需要字符串中的字符，而不需要操作下标。

但对is\_abecedarian函数我们需要比较相邻的字母，使用for循环比较困难：

```
def is_abecedarian(word):
    previous = word[0]
    for c in word:
        if c < previous:
            return False
        previous = c
    return True
```

或者也可以使用递归：

```
def is_abecedarian(word):
    if len(word) <= 1:
        return True
    if word[0] > word[1]:
        return False
    return is_abecedarian(word[1:])
```

还有一个办法是使用while循环：

```
def is_abecedarian(word):
    i = 0
    while i < len(word)-1:
        if word[i+1] < word[i]:
            return False
        i = i+1
    return True
```

循环开始于*i*=0并当*i*=*len(word)*-1时结束。每次迭代时，比较第*i*个字符（可以看成是当前字符）和第*i*+1个字符（可以看成是下一个字符）。

如果下一个字符比当前字符小（按照字母顺序在前），则我们发现了一个破坏字母顺序的断点，可以返回False。

如果我们没有找到任何断点而结束循环，则这个单词通过了测试。为了说服自己循环是正确结束的，可以考虑像‘flossy’这样的例子。这个单词的长度是6，所以最后一次循环时i是4，即是倒数第二个字符的下标。在最后一个循环中，会比较倒数第二个和最后一个字符，这正是我们所期待的。

下面是is\_palindrome函数（参考练习 6-6）的一个版本，它使用两个下标；一个从0开始递增；另一个从最后开始递减。

```
def is_palindrome(word):
    i = 0
    j = len(word)-1

    while i < j:
        if word[i] != word[j]:
            return False
        i = i+1
        j = j-1

    return True
```

或者，如果你发现这是一个已经解决的问题的特例的话，你可能会这么写：

```
def is_palindrome(word):
    return is_reverse(word, word)
```

前提是假定你已经做过了练习8-9。

## 9.5 调试

测试程序很难。本章中的函数相对容易测试，因为你可以简单地手动验证结果。即便如此，要选择一组可以测试到所有可能的错误的单词，也是很困难，甚至不可能的。

举`has_no_e`作为例子，有两个很明显的用例可以检测：包含‘e’的单词应该返回`False`；不包含‘e’的应当返回`True`。为这两种情况找到具体的单词没有问题。

但对每种情况来说，也存在一些不那么明显的具体情况。在所有包含‘e’的单词中，你应当测试以‘e’开头的单词，也应当测试以‘e’结尾，以及‘e’在单词中部的情况。你应当测试长单词、短单词及非常短的单词，比如空字符串。空字符串是**特殊情形**（special case）的一个例子。特殊情形往往不那么明显，但又常常隐藏着错误。

除了自己生成的测试用例之外，还可以使用类似`words.txt`这样的单词表来测试你的程序。通过扫描输出，可能会发现错误，但请注意：你可能发现一种类型的错误（不应该被包含但却被包含的单词），但对另一种类型的则不能发现（应该被包含，但却没有出现的单词）。

总之，测试可以帮助你发现bug，但生成一组好的测试用例不容易。并且即使有好的测试用例，也不能确定程序是完全正确的。

引用一个传奇计算机科学家的话：

程序测试可以用来显示bug的存在，但无法显示它们的缺席！  
(Program testing can be used to show the presence of bugs,  
but never to show their absence!)

——Edsger W. Dijkstra

## 9.6 术语表

文件对象（file object）：用来表示一个打开的文件的值。

问题识别（problem recognition）：解决问题的一种方式：把问题表述为已经解决的某个问题的特例。

特殊情形（special case）：一种不典型或者不明显（因此更可能没有正确处理）的测试用例。

## 9.7 练习

### 练习9-7

本练习中的问题是基于广播节目《车迷天下》（Car Talk）中出现的一个谜题而设计的

（<http://www.cartalk.com/content/puzzler/transcripts/200725>）：  
：

给我一个包含3组连续的成对字母的单词。我会给你几个几乎可以达到要求，却还差一点的词作为例子。例如，单词committee，即c-o-m-m-i-t-t-e-e。这个单词除了‘i’不满足条件外，会是一个好例子。或者Mississippi：M-i-s-s-i-s-s-i-p-p-i。如果你能够拿掉其中的i，则它也符合要求。但确实有这么一个单词，并且就我所知，它可能是满足这个条件的唯一的单词。当然也可能存在500个，但我只能想到一个。它是什么呢？

编写一个程序来找到它。解答：  
<http://thinkpython.com/code/cartalk1.py>。

### 练习9-8

下面是另一个《车迷天下》中的谜题  
(<http://www.cartalk.com/content/puzzler/transcripts/200803>)：

“有一天我正在高速公路上开车，碰巧注意到里程表。和大部分里程表一样，它显示6位整数的英里数。所以，例如我的车有300 000英里里程，则会看到3-0-0-0-0-0。

“那天我看到的里程数很有意思。我发现最后4位数是回文的；也就是说，它们不论是正序还是逆序地看都一样。例如，5-4-4-5是一个回文，所以我的里程表可能显示为3-1-5-4-4-5。

“1英里之后，后5位数组成一个回文。例如，它可以是3-6-5-4-5-6。再过1英里，6位数的中间4位是一个回文。而接下来，你准备好了吗？又1英里过去，所有的6位数都成了回文！

“问题是，我第一次看里程表时，它的示数是多少？”

编写一个Python程序，检测全部的6位数，并打印出可以满足上面这些要求的数字。解答：

<http://thinkpython.com/code/cartalk2.py>。

### 练习9-9

下面是另一个《车迷天下》的谜题，你可以使用一个搜索来解决（<http://www.cartalk.com/content/puzzler/transcripts/200813>）：

“最近我访问母亲家，并发现我的年龄的两位数正好是她的年龄的两位数的倒序。例如，如果她是73岁，我是37岁。我们好奇这种事情这些年来发生过几次，但很快我们的话题就偏转到其他地方，所以没有得到答案。

“当我回家后，发现我们的年龄互为倒序的事情至今为止发生过6次。我也发现我们如果顺利的话接下来几年还会再遇到一次，并且在那之后如果我们真的很幸运，还能再遇到一次。换句话说，它总共可能发生8次。所以问题是，我现在年龄多大？”

编写一个Python程序，为这个谜题搜索答案。提示：你可能会发现字符串方法`zfill`有用。

解答：<http://thinkpython.com/code/cartalk3.py>。

---

① 作者在上一段话中模仿了Gadsby的风格，不使用字母“e”，所以说话的风格很怪。因此到这里，他就说“All right, I’ll stop now”，意思是停止这种怪异风格的描述。但这个意思无法在译文中表达。——译者注

# 第10章 列表

## 10.1 列表是一个序列

和字符串相似，列表（list）是值的序列。在字符串中，这些值是字符；在列表中，它可以是任何类型。列表中的值称为元素（element），有时也称为列表项（item）。

创建一个列表有好几种方式：最简单的方式是使用方括号（[ 与 ]）将元素括起来。

```
[10, 20, 30, 40]  
['crunchy frog', 'ram bladder', 'lark vomit']
```

第一个例子是4个整数的列表。第二个例子是3个字符串的列表。列表中的元素并不一定非得是同一类型的。下面的列表包含了一个字符串、一个浮点数、一个整数及（瞧！）另一个列表：

```
['spam', 2.0, 5, [10, 20]]
```

列表中出现的列表是嵌套的（nested）。

不包含任何元素的列表称为空列表；你可以使用空方括号[]来创建空列表。

如你所预料的，列表可以赋值给变量：

```
>>> cheeses = ['Cheddar', 'Edam', 'Gouda']  
>>> numbers = [17, 123]  
>>> empty = []  
>>> print cheeses, numbers, empty  
['Cheddar', 'Edam', 'Gouda'] [17, 123] []
```

## 10.2 列表是可变的

访问列表元素的语法和访问字符串中字符的语法是一样的——使用方括号操作符。方括号中的表达式指定下标。请记得下标是从0开始的：

```
>>> print cheeses[0]
Cheddar
```

和字符串不同的是，列表是可变的。当方括号操作符出现在赋值语句的左侧时，它用于指定列表中哪个元素会被赋值。

```
>>> numbers = [17, 123]
>>> numbers[1] = 5
>>> print numbers
[17, 5]
```

numbers的第1位元素，原先的值是123，现在是5了。

你可以把列表看作下标和元素之间的关联。这种关联称为映射(mapping)；每个下标“映射到”元素中的一个。图10-1显示了cheeses, numbers和empty的状态图。

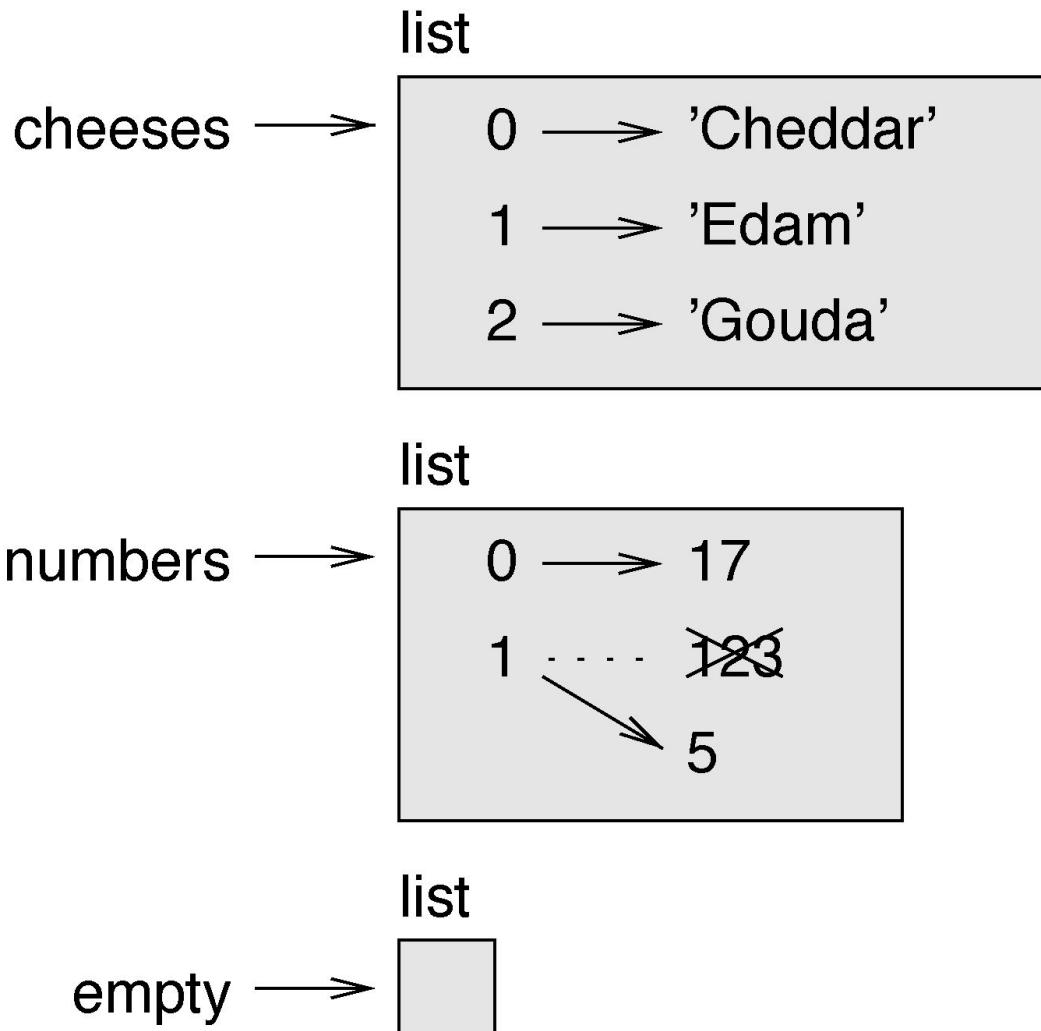


图10-1 状态图

在图10-1中，外面写有“list”的图框表示列表，里面显示的是列表中的元素。`cheeses`变量引用着一个列表，包含3个元素，下标分别是0, 1和2。`numbers`包含两个元素；本图显示了其第二个元素从123重新赋值为5的过程。`empty`引用一个没有任何元素的空列表。

列表下标和字符串下标工作方式相同。

- 任何整型的表达式都可以用作下标。
- 如果尝试读写一个并不存在的元素，则会得到`IndexError`。
- 如果下标是负数，则从列表的结尾处反过来数下标访问。

`in`操作符也可以用于列表。

```
>>> cheeses = ['Cheddar', 'Edam', 'Gouda']
>>> 'Edam' in cheeses
True
>>> 'Brie' in cheeses
False
```

## 10.3 遍历一个列表

遍历一个列表元素的最常见方式是使用`for`循环。语法和字符串的遍历相同：

```
for cheese in cheeses:  
    print cheese
```

当你只需要读取列表的元素本身时，这样的遍历方式很好。但如果需要写入或者更新元素时，则需要下标。一个常见的方式是使用函数`range`和`len`：

```
for i in range(len(numbers)):  
    numbers[i] = numbers[i] * 2
```

这个循环遍历列表，并更新每个元素。`len`返回列表中元素的个数。`range`返回一个下标的列表，从0到 $n-1$ ，其中 $n$ 是列表的长度。每次迭代时，`i`获得下一个元素的下标。循环体中的赋值语句使用`i`来读取元素的旧值并赋值为新值。

在空列表上使用`for`循环，则循环体从不会被执行：

```
for x in []:  
    print 'This never happens.'
```

虽然列表可以包含其他的列表，嵌套的列表仍然被看作一个单独的元素。下面的列表长度是4：

```
['spam', 1, ['Brie', 'Roquefort', 'Poul le Veq'], [1, 2, 3]]
```

## 10.4 列表操作

+操作符可以拼接列表：

```
>>> a = [1, 2, 3]
>>> b = [4, 5, 6]
>>> c = a + b
>>> print c
[1, 2, 3, 4, 5, 6]
```

相似地，\*操作符重复一个列表多次：

```
>>> [0] * 4
[0, 0, 0, 0]
>>> [1, 2, 3] * 3
[1, 2, 3, 1, 2, 3, 1, 2, 3]
```

第一个例子重复列表[0]四次。第二个例子重复列表[1, 2, 3]三次。

## 10.5 列表切片

切片操作符也可以用于列表：

```
>>> t = ['a', 'b', 'c', 'd', 'e', 'f']
>>> t[1:3]
['b', 'c']
>>> t[:4]
['a', 'b', 'c', 'd']
>>> t[3:]
['d', 'e', 'f']
```

如果省略掉第一个下标，则切片从列表开头开始。如果省略掉第二个下标，则切片至列表结尾结束。如果两个下标都省略，则切片就是整个列表的副本。

```
>>> t[:]
['a', 'b', 'c', 'd', 'e', 'f']
```

因为列表是可变的，所以在对列表进行折叠、拉伸或者破坏操作之前，复制一份是很有用的。

如果切片操作符出现在赋值语句的左侧，则可以更新多个元素：

```
>>> t = ['a', 'b', 'c', 'd', 'e', 'f']
>>> t[1:3] = ['x', 'y']
>>> print t
['a', 'x', 'y', 'd', 'e', 'f']
```

## 10.6 列表方法

Python为列表提供了不少操作方法。比如，`append`可以在列表尾部添加新的元素：

```
>>> t = ['a', 'b', 'c']
>>> t.append('d')
>>> print t
['a', 'b', 'c', 'd']
```

`extend`方法接收一个列表作为参数，并将其所有的元素附加到列表中：

```
>>> t1 = ['a', 'b', 'c']
>>> t2 = ['d', 'e']
>>> t1.extend(t2)
>>> print t1
['a', 'b', 'c', 'd', 'e']
```

这个例子中`t2`没有被修改。

`sort`方法将列表中的元素从低到高重新排列：

```
>>> t = ['d', 'c', 'e', 'b', 'a']
>>> t.sort()
>>> print t
['a', 'b', 'c', 'd', 'e']
```

列表的方法全都是无返回值的；它们修改列表，并返回`None`。如果你不小心写了`t = t.sort()`，则很可能会得到失望的结果。

## 10.7 映射、过滤和化简

如果想把列表中所有的元素加起来，你可以使用下面这样的循环：

```
def add_all(t):
    total = 0
    for x in t:
        total += x
    return total
```

total被初始化为0。每次循环中，x获取列表中的一个元素。+=操作符为更新变量提供了一个简洁的方式。这个增加赋值语句：

```
total += x
```

等价于：

```
total = total + x
```

随着循环的执行，total会累积列表中的值的和；这样使用一个变量有时称为累加器（accumulator）。

对列表元素累加是如此常见的操作，以至于Python提供了一个内置函数sum：

```
>>> t = [1, 2, 3]
>>> sum(t)
6
```

类似这样，将一个序列的元素值合起来到一个单独的变量的操作，有时称为化简（reduce）。

练习10-1

编写一个函数nested\_sum，接收一个由内嵌的整数列表组成的列表作为形参，并将内嵌列表中的值全部加起来。

有时候你想要在遍历一个列表的同时构建另一个列表。例如，下面的函数接收一个字符串列表，并返回一个新列表，其元素是大写的字符串：

```
def capitalize_all(t):
    res = []
    for s in t:
        res.append(s.capitalize())
    return res
```

res初始化为一个空列表；每次循环，我们给它附加一个元素。所以res也是一种累加器。

像capitalize\_all这样的操作，有时被称为映射（map），因为它将一个函数（在这个例子里是capitalize方法）“映射”到一个序列的每个元素上。

### 练习10-2

使用capitalize\_all来编写一个函数capitalize\_nested，接收一个由内嵌的字符串列表组成的列表作为形参，并返回一个新列表，其元素为内嵌的大写字符串的列表。

另一个常见的操作是选择列表中的某些元素，并返回一个子列表。例如，下面的函数接收一个字符串列表，并返回那些只包含大写字母的字符串：

```
def only_upper(t):
    res = []
    for s in t:
        if s.isupper():
            res.append(s)
    return res
```

isupper是一个字符串方法，当字符串中只包含大写字母时返回True。

类似only\_upper这样的操作称为过滤（filter），因为它选择列表中的某些元素，并过滤掉其他的元素。

列表的绝大多数常用操作都可以用映射、过滤和化简的组合来表达。因为这些操作非常常见，Python提供了语言特性直接支持它们，包括内置函数map，以及被称为“列表推导”（list comprehension）的操作符。

### 练习10-3

编写一个函数，接收一个数字列表，并返回其累积和，即一个新的列表，其第  $i$  位元素是原始列表的前  $i+1$  个元素的和。例如， $[1, 2, 3]$  的累积和是  $[1, 3, 6]$ 。

## 10.8 删 除 元 素

从列表中删除元素，有多种方法。如果你知道元素的下标，可以使用`pop`:

```
>>> t = ['a', 'b', 'c']
>>> x = t.pop(1)
>>> print t
['a', 'c']
>>> print x
b
```

`pop`修改列表，并返回被删除掉的值。如果不提供下标，它会删除并返回最后一个元素。

如果你不需要使用删除的值，可以使用`del`操作符:

```
>>> t = ['a', 'b', 'c']
>>> del t[1]
>>> print t
['a', 'c']
```

如果你知道要删除的元素（而不是下标），则可以使用`remove`:

```
>>> t = ['a', 'b', 'c']
>>> t.remove('b')
>>> print t
['a', 'c']
```

`remove`方法的返回值是`None`。

若要删除多个元素，可以使用`del`和切片下标:

```
>>> t = ['a', 'b', 'c', 'd', 'e', 'f']
>>> del t[1:5]
>>> print t
['a', 'f']
```

和通常一样，切片会从第一个下标开始，选择所有的元素，直到第二个下标（并不包含）。

练习10-4

编写一个函数middle，接收一个列表作为形参，并返回一个新列表，包含除了第一个和最后一个元素之外的所有元素。所以middle([1, 2, 3, 4])应该返回[2, 3]。

#### 练习10-5

编写一个函数chop，接收一个列表，修改它，删除它的第一个和最后一个元素，并返回None。

## 10.9 列表和字符串

字符串是字符的序列，而列表是值的序列，但字符的列表和字符串并不相同。若要将一个字符串转换为一个字符的列表，可以使用函数list：

```
>>> s = 'spam'  
>>> t = list(s)  
>>> print t  
['s', 'p', 'a', 'm']
```

因为list是内置函数的名称，所以应当尽量避免使用它作为变量名称。我也避免使用l，因为它看起来太像数字1了。因而我使用t。

list函数会将字符串拆成单个的字母。如果想要将字符串拆成单词，可以使用split方法：

```
>>> s = 'pining for the fjords'  
>>> t = s.split()  
>>> print t  
['pining', 'for', 'the', 'fjords']
```

split还接收一个可选的形参，称为分隔符（delimiter），用于指定用哪个字符来分隔单词。下面的例子中使用连字符（-）作为分隔符：

```
>>> s = 'spam-spam-spam'  
>>> delimiter = '-'  
>>> s.split(delimiter)  
['spam', 'spam', 'spam']
```

join是split的逆操作。它接收字符串列表，并拼接每个元素。join是字符串的方法，所以你必须在分隔符上调用它，并传入列表作为实参：

```
>>> t = ['pining', 'for', 'the', 'fjords']  
>>> delimiter = ' '  
>>> delimiter.join(t)  
'pining for the fjords'
```

在这个例子里，分隔符是空格，所以join会在每个单词之间放一个空格。若想不用空格直接连接字符串，可以使用空字符串，''，作为分隔符。

## 10.10 对象和值

如果我们执行下面的赋值语句：

```
a = 'banana'  
b = 'banana'
```

我们知道a和b都是一个字符串的引用。但我们不知道它们是否指向同一个字符串。有两种可能的状态，如图10-2所示。

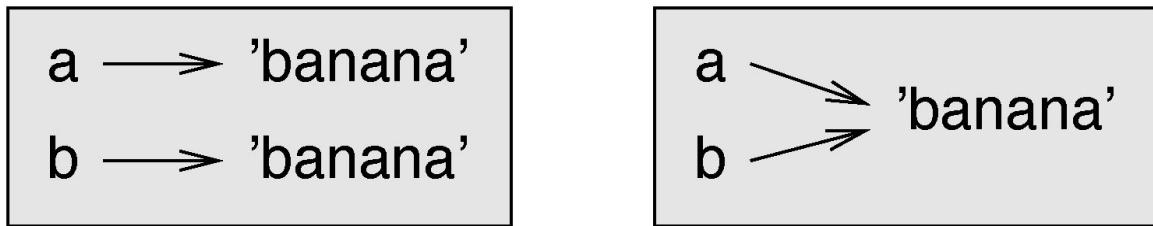


图10-2 状态图

一种可能是，a和b引用着不同的对象，它们的值相同。另一种情况下，它们指向同一个对象。

要检查两个变量是否引用同一个对象，可以使用`is`操作符。

```
>>> a = 'banana'  
>>> b = 'banana'  
>>> a is b  
True
```

在这个例子里，Python只建立了一个字符串对象，而a和b都引用它。

但当你新建两个列表时，会得到两个对象：

```
>>> a = [1, 2, 3]  
>>> b = [1, 2, 3]  
>>> a is b  
False
```

所以状态图如图10-3所示。

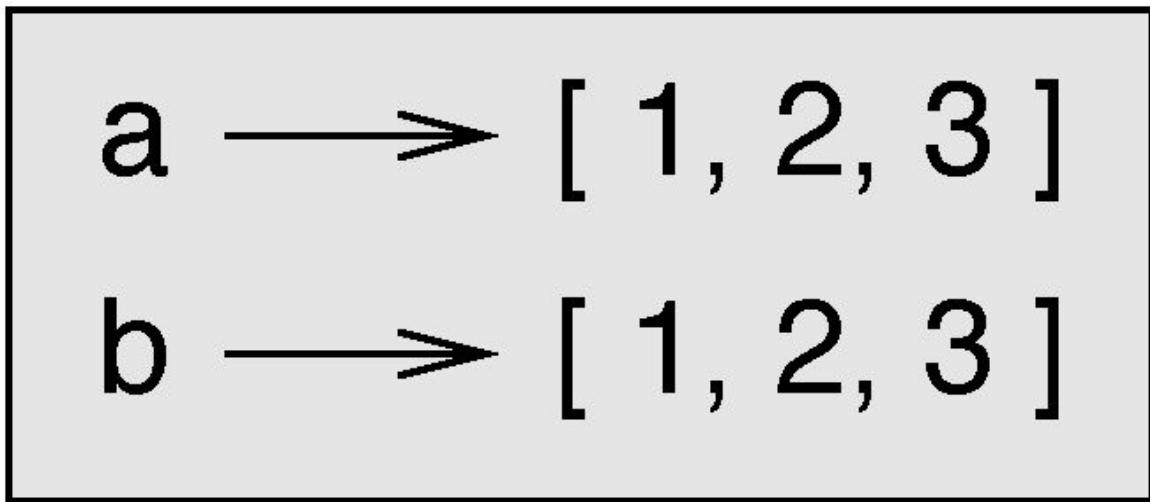


图10-3 状态图

在这个例子里我们会说这两个列表是相等的 (equivalent)，因为它们有相同的元素，但它们不是相同的 (identical)，因为它们并不是同一个对象。如果两个对象相同，则必然也相等，但如果两个对象相等，并不一定相同。

至今为止，我们都不加区分地使用“对象”和“值”，但更精确的说法是对象有一个值。如果你运行`[1, 2, 3]`，会得到一个列表对象，它的值是一个整数的序列。如果另一个列表包含相同的元素，我们说它有相同的值，但它们不是同一个对象。

## 10.11 别名

如果a引用一个对象，而你赋值b = a，则两个变量都会引用同一个对象：

```
>>> a = [1, 2, 3]
>>> b = a
>>> b is a
True
```

这里的状态图如图10-4所示。

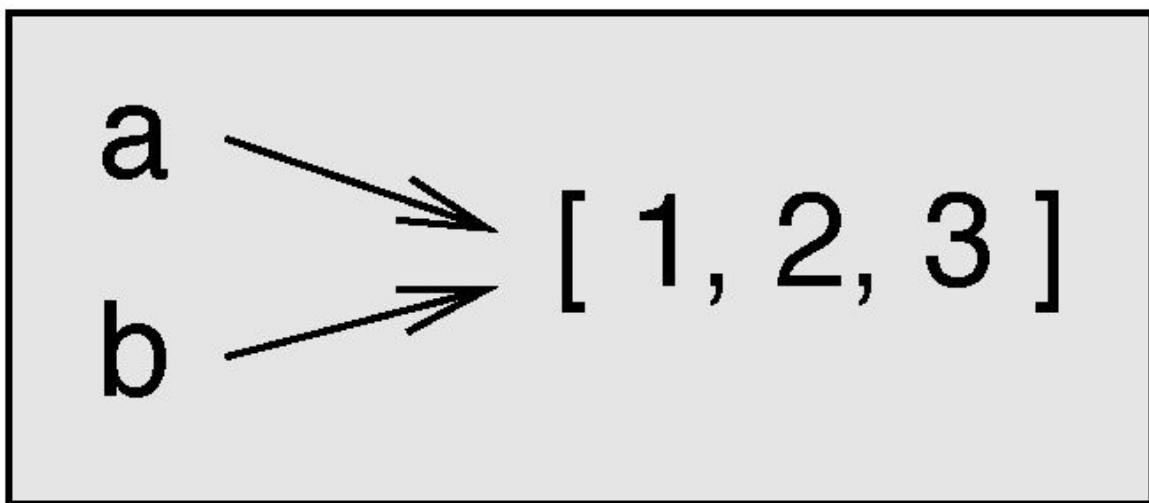


图10-4 状态图

变量和对象之间的关联关系称为引用（reference）。在这个例子里，有两个指向同一对象的引用。

当一个对象有多个引用，并且引用有不同的名称时，我们说这个对象有别名（alias）。

如果有别名的对象是可变的，则对一个别名的修改会影响另一个：

```
>>> b[0] = 17
>>> print a
[17, 2, 3]
```

虽然这种行为可能很有用，但它也容易导致错误。通常来说，当处理可变对象时，避免使用别名会更加安全。

对于字符串这样的不可变对象，别名则不会带来问题。在下面的例子中：

```
a = 'banana'  
b = 'banana'
```

不论a和b是否引用同一个字符串，都不会有什么区别。

## 10.12 列表参数

当你将一个列表作为参数传入到函数中，函数会得到列表的一个引用。如果函数中修改了列表参数，则调用者也会看到这个修改。例如，`delete_head`函数删除列表中的第一个元素：

```
def delete_head(t):
    del t[0]
```

下面使用它：

```
>>> letters = ['a', 'b', 'c']
>>> delete_head(letters)
>>> print letters
['b', 'c']
```

参数`t`和变量`letters`是同一个对象的别名。栈图如图 10-5所示。

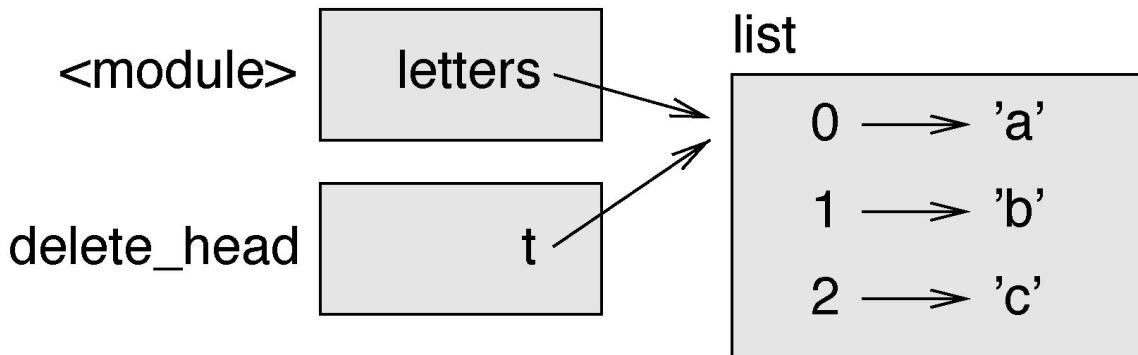


图10-5 栈图

因为列表被两个帧共享，所以我将它画在中间。

区分修改列表的操作和新建列表的操作十分重要。例如，`append`方法修改列表，但是`+`操作符新建一个列表：

```
>>> t1 = [1, 2]
>>> t2 = t1.append(3)
>>> print t1
[1, 2, 3]
>>> print t2
None
```

```
>>> t3 = t1 + [4]
>>> print t3
[1, 2, 3, 4]
```

这个区别，在你编写希望修改列表的函数时十分重要。例如，下面的函数并不会删除列表的开头：

```
def bad_delete_head(t):
    t = t[1:] # 错！
```

切片操作会新建一个列表，而赋值操作会让`t`引用指向这个新的列表，但这些操作对作为参数传入的那个列表没有任何影响。

另外一种方法是编写函数创建和返回一个新的列表。例如，`tail`返回除了第一个以外所有的元素的列表：

```
def tail(t):
    return t[1:]
```

这个函数不会修改原始列表。下面的代码展示如何使用它：

```
>>> letters = ['a', 'b', 'c']
>>> rest = tail(letters)
>>> print rest
['b', 'c']
```

## 10.13 调试

对列表（以及其他可变对象）的不慎使用，可能会导致长时间的调试。下面介绍一些常见的陷阱，以及如何避免它们。

1. 别忘了大部分列表方法都是修改参数并返回None的。这和字符串的方法正相反，字符串方法新建一个字符串，并留着原始的字符串不动。

如果你习惯于写下面这样的字符串代码：

```
word = word.strip()
```

则容易倾向于这么写列表代码：

```
t = t.sort()          # 错！
```

因为sort返回None，接下来对t进行的操作很可能会失败。

在使用列表方法和操作符之前，应当仔细阅读文档，并在交互模式中测试它们。列表和其他序列（如字符串）共有的方法和操作符的文档在<http://docs.python.org/lib/typesseq.html>中。而只对可变序列有效的方法和操作符文档在<http://docs.python.org/lib/typesseq-mutable.html>中。

2. 选择一种风格，并坚持不变。

列表的问题之一是同样的事情有太多种可用的做法。例如，要从列表中删除一个元素，你可以使用pop、remove、del或者甚至是切片赋值。

要添加一个元素，可以使用append方法或者+操作符。假设t是一个列表，x是一个列表元素，下面的操作是正确的：

```
t.append(x)  
t = t + [x]
```

而下面的操作是错误的：

```
t.append([x])      # 错!  
t = t.append(x)    # 错!
```

```
t + [x]          # 错!
t = t + x      # 错!
```

在交互模式中试验这些例子，确保你明白它们的运行细节。注意只有最后一个会导致运行时错误；其他的3个都是合法的，但是它们的结果不正确。

### 3. 通过复制来避免别名。

如果你想要使用类似`sort`的方法来修改参数，但也需要保留原先的列表，可以复制一个副本：

```
orig = t[:]
t.sort()
```

在这个例子里也可以使用内置函数`sorted`，它会返回一个新的排好序的列表，并且留着原先的列表不动。但这样你也需要避免使用`sorted`作为变量名！

## 10.14 术语表

列表 (list) : 值的序列。

元素 (element) : 列表 (或其他序列) 中的一个值，也称为列表项。

下标 (index) : 标明元素在列表中的位置的整数值。

嵌套列表 (nested list) : 作为其他列表的元素的列表。

列表遍历 (list traversal) : 按顺序遍历访问列表中的每个元素。

映射 (mapping) : 一种关联，表示一个集合的每个元素和另一个集合的元素如何对应。例如，列表是从下标到元素的映射。

累加器 (accumulator) : 在循环中用于加和或者累积某个结果的变量。

增加赋值 (augmented assignment) : 使用类似`+ =`操作符来更新变量值的语句。

化简 (reduce) : 一种处理模式，遍历一个序列，并将元素的值累积起来计算为一个单独的结果。

映射 (map) : 一种处理模式，遍历一个序列，对每个元素进行操作。

过滤 (filter) : 一种处理模式，遍历列表，并选择满足某种条件的元素。

对象 (object) : 变量可以引用的东西。对象有类型和值。

相等 (equivalent) : 拥有相同的值。

相同 (identical) : 是同一个对象 (并且也意味着相等)。

引用 (reference) : 变量和它的值之间的关联。

别名 (aliasing) : 多个变量同时引用一个对象的情况。

分隔符 (delimiter) : 用于分隔字符串的一个字符或字符串。

## 10.15 练习

### 练习10-6

编写一个函数`is_sorted`, 接收一个列表作为形参, 并当列表是按照升序排好序的时候返回`True`, 否则返回`False`。你可以假定(作为前置条件)列表的元素是可以使用关系操作符`<`, `>`等比较的。

例如, `is_sorted([1, 2, 2])`应当返回`True`, 而`is_sorted(['b', 'a'])`应当返回`False`。

### 练习10-7

两个单词, 如果重新排列其中一个的字母可以得到另一个, 它们互为回文(anagram)。编写一个函数`is_anagram`, 接收两个字符串, 当它们互为回文时返回`True`。

### 练习10-8

(所谓的)生日悖论:

1. 编写一个函数`has_duplicates`接收一个列表, 当其中任何一个元素出现多于一次时返回`True`。它不应当修改原始列表;
2. 如果你的班级中有23个学生, 那么其中有两人生日相同的几率有多大? 你可以通过随机生成23个生日的样本并检查是否有相同的匹配来估计这个几率。提示: 可以使用`random`模块中的`randint`函数来生成随机生日。

你可以在[http://en.wikipedia.org/wiki/Birthday\\_paradox](http://en.wikipedia.org/wiki/Birthday_paradox)阅读这个问题的资料, 解答可以从<http://thinkpython.com/code/birthday.py>下载。

### 练习10-9

编写一个函数`remove_duplicates`, 接收一个列表, 并返回一个新列表, 其中只包含原始列表的每个元素的唯一一份。提示: 它们不需要顺序相同。

### 练习10-10

编写一个函数，读取文件words.txt，并构建一个列表，每个元素是一个单词。给这个函数编写两个版本，其中一个使用append方法，另一个使用`t = t + [x]`的用法。哪一个运行时间更长？为什么？

提示：使用time模块来计算花费的时间。解答：  
<http://thinkpython.com/code/wordlist.py>。

#### 练习10-11

要检查一个单词是否出现在单词列表中，你可以使用in操作符，但由于它需要按顺序搜索所有单词，可能会比较慢。

因为单词是按字母顺序排列的，我们可以使用二分查找（也叫做二分搜索）来加快速度。二分查找的过程类似于在字典中查找单词。你从中间开始，检查需要找的单词是不是在列表中间出现的单词之前，如果是，则继续用同样的方法搜索前半部分。否则搜索后半部分。

不论何种情况，都将搜索空间减小了一半。如果单词列表有113,809个单词，那么大概会耗费17步来找到单词，或者确定它不在列表之中。

编写一个函数bisect，接收一个排好序的列表，以及一个目标值，当目标值在列表之中，返回其下标，否则返回None。

或者你可以阅读bisect模块的文档，并使用它！ 解答：  
<http://thinkpython.com/code/inlist.py>。

#### 练习10-12

两个单词，如果其中一个是另一个的反向序列，则称它们为“反向对”。编写一个程序找到单词表中出现的全部反向对。解答：

[http://thinkpython.com/code/reverse\\_pair.py](http://thinkpython.com/code/reverse_pair.py)。

#### 练习10-13

两个单词，如果从每个单词中交错取出一个字母可以组成一个新的单词，我们称它们为“互锁”（interlocking）。例如，“shoe”和“cold”可以互锁组成单词“schooled”。解答：  
<http://thinkpython.com/code/interlock.py>。鸣谢：这个练习启发自<http://puzzlers.org>的一个示例。

1. 编写一个程序找到所有互锁的词。提示：不要穷举所有的词对！

2. 能不能找到“三互锁”的单词？也就是，从第一、第二或者第三个字母开始，每第三个字母合起来可以形成一个单词。

# 第11章 字典

字典类似于列表，但更加通用。在列表中，下标必须是整数；而在字典中，下标（几乎）可以是任意类型。

你可以把字典看作下标（称为键）集合与值集合之间的映射。每个键都映射到一个值上。键和值之间的关联被称为键值对（key-value pair），或者有时称为一项（item）。

作为示例，我们构建一个字典，将英语单词映射到西班牙语上，所以键和值的类型都是字符串。

函数`dict`新建一个不包含任何项的字典。因为`dict`是内置函数的名称，应当避免使用它作为变量名。

```
>>> eng2sp = dict()
>>> print eng2sp
{}
```

这里花括号`{}`表示一个空的字典。想要给字典添加新项，可以使用方括号操作符：

```
>>> eng2sp['one'] = 'uno'
```

这一行代码创建一个新项，将键`'one'`映射到值`'uno'`上。如果我们再次打印这个字典，可以看到一个键值对，以冒号分隔：

```
>>> print eng2sp
{'one': 'uno'}
```

这种输出格式也同样是输入的格式。例如，你可以创建一个包含3项的新字典：

```
>>> eng2sp = {'one': 'uno', 'two': 'dos', 'three': 'tres'}
```

但如果你打印`eng2sp`，可能会觉得奇怪：

```
>>> print eng2sp
{'one': 'uno', 'three': 'tres', 'two': 'dos'}
```

字典中键值对的顺序并不相同。实际上，如果你在自己的电脑上输入相同的示例，可能会得到另一个不同的结果。总之，字典中各项的顺序是不可预料的。

但这并不是问题，因为字典的元素从来不使用整数下标进行查找。相对地，它使用键来查找对应的值：

```
>>> print eng2sp['two']
'dos'
```

如果键'two'总是映射到值'dos'上，那么各项的顺序其实并不重要。

如果一个键并不在字典之中，你会得到一个异常：

```
>>> print eng2sp['four']
KeyError: 'four'
```

`len`函数可以用在字典上，它返回键值对的数量：

```
>>> len(eng2sp)
3
```

`in`操作符也可以用在字典上，它告诉你一个值是不是字典中的键（是字典中的值则不算）。

```
>>> 'one' in eng2sp
True
>>> 'uno' in eng2sp
False
```

若要查看一个值是不是出现在字典的值中，可以使用方法`values`，它会以列表的形式返回字典所有的值，并可以应用`in`操作符：

```
>>> vals = eng2sp.values()
>>> 'uno' in vals
True
```

`in`操作符对列表和字典使用不同的算法实现。对于列表，它使用搜索算法，如8.6节所示。当列表变长时，搜索时间会随之变长。而对于字典，Python使用一个称为散列表（hashtable）的算法。它有一个值得注意的特点：不管字典中有多少项，`in`操作符花费的时间都差不多。这里我不解释

其中的原因，但你可以在[http://en.wikipedia.org/wiki/Hash\\_table](http://en.wikipedia.org/wiki/Hash_table)阅读到更多的相关信息。

### 练习11-1

编写一个函数，读入words.txt的单词列表，并将它们作为键保存到一个字典中。字典的值是什么并不重要。然后你就可以使用in操作符来快速地检查一个字符串是否在这个字典之中。

如果你做过了练习10-11，可以将这个实现与列表的in操作符以及二分查找，进行速度的对比。

## 11.1 使用字典作为计数器集合

假设给定一个字符串，你想要计算每个字母出现的次数。有几种可能的实现方法：

1. 你可以创建26个变量，每个变量对应字母表上的一个字母。接着遍历字符串，对每一个字符，增加对应的计数器。你可能需要使用一个链式条件判断。
2. 你可以创建一个包含26个元素的列表。接着可以将每个字符转换为一个数字（使用内置函数`ord`），使用这个数字作为列表的下标，并增加对应的计数器。
3. 你可以建立一个字典，以字符作为键，以计数器作为相应的值。第一次遇到某个字符时，在字典中添加对应的项。之后可以增加一个已经存在的项的值。

这几种方案进行相同的计算，但实现计算的方式不一样。

实现（implementation）是进行某种计算的一个具体方式；有的实现比其他的更好。例如，字典实现的优势之一是我们并不需要预先知道字符串中可能出现哪些字母，因而只需为真正出现过的字母分配空间。

下面是这个实现的代码：

```
def histogram(s):
    d = dict()
    for c in s:
        if c not in d:
            d[c] = 1
        else:
            d[c] += 1
    return d
```

这个函数的名称是直方图（histogram），它是一个统计学术语，表示一个计数器（或者说频率）的集合。

函数的第一行创建一个空的字典。`for`循环遍历字符串。每次迭代中，如果字符`c`不在字典中，我们就创建一个新项，其键是`c`，其值初始化为

1（因为我们已经见到这个字符一次了）。如果c已经在字典之中，我们增加d[c]。

下面是这个函数的使用方式：

```
>>> h = histogram('brontosaurus')
>>> print h
{'a': 1, 'b': 1, 'o': 2, 'n': 1, 's': 2, 'r': 2, 'u': 2, 't': 1}
```

这个直方图显示，字母'a'和'b'出现了1次；'o'出现了两次，依此类推。

### 练习11-2

字典有一个方法get，接收一个键以及一个默认值。如果键出现在字典中，get返回对应的值；否则它返回默认值。例如：

```
>>> h = histogram('a')
>>> print h
{'a': 1}
>>> h.get('a', 0)
1
>>> h.get('b', 0)
0
```

使用get将histogram写得更紧凑一些。你应当可以消除掉if语句。

## 11.2 循环和字典

如果在for循环中使用字典，会遍历字典的键。例如，print\_hist函数打印字典的每一个键以及对应的值：

```
def print_hist(h):
    for c in h:
        print c, h[c]
```

下面是这个函数输出的样子：

```
>>> h = histogram('parrot')
>>> print_hist(h)
a 1
p 1
r 2
t 1
o 1
```

同样地，键的出现没有特定的顺序。

### 练习11-3

字典有一个方法keys，以列表的方式返回字典所有的键，并不保证特定的顺序。

修改print\_hist按照字典顺序打印字典的键和它们的值。

## 11.3 反向查找

给定一个字典`d`和键`k`, 找到对应的值`v = d[k]`非常容易。这个操作称为查找（lookup）。

但是如果你有`v`, 而想找到`k`时怎么办? 这里有两个问题: 首先, 可能存在多个键映射到同一个值`v`上。随不同的应用场景, 你也许可以挑其中一个, 或者也许需要建立一个列表来保存所有的键。其次, 并没有可以进行反向查找的简单语法; 你需要使用搜索。

下面是一个函数, 接收一个值, 并返回映射到该值的第一个键:

```
def reverse_lookup(d, v):
    for k in d:
        if d[k] == v:
            return k
    raise ValueError
```

这个函数是搜索模式的又一个实例。但它使用了一个我们还没见过的语言特性, `raise`语句。`raise`语句会生成一个异常; 在这个例子里它生成一个`ValueError`, 通常用来表示参数的值有问题。

如果我们到达了循环的结尾, 就意味着`v`在字典中没有作为值出现过, 所以我们抛出一个异常。

下面的例子展示了一个成功的反向查找:

```
>>> h = histogram('parrot')
>>> k = reverse_lookup(h, 2)
>>> print k
r
```

以及一个不成功的反向查找:

```
>>> k = reverse_lookup(h, 3)
Traceback (most recent call last):
  File "<stdin>", line 1, in ?
  File "<stdin>", line 5, in reverse_lookup
ValueError
```

当你自己抛出异常时，效果和Python抛出异常是一样的：它会打印出一个回溯和一个错误信息。

`raise`语句也接收一个可选的参数用来详细描述错误。例如：

```
>>> raise ValueError, 'value does not appear in the dictionary'  
Traceback (most recent call last):  
  File "<stdin>", line 1, in ?  
ValueError: value does not appear in the dictionary
```

反向查找远远慢于正向查找；如果你频繁这么做，或者字典非常大时，会对程序的性能有很大影响。

#### 练习11-4

修改`reverse_lookup`函数，使得它建立并返回一个包含所有映射到`v`的键的列表，或者，在没有找到对应键时返回空列表。

## 11.4 字典和列表

列表可以在字典中以值的形式出现。例如，如果你遇到一个将字母映射到频率的字典，可能会想要反转它；也就是说，建立一个字典，将频率映射到字母上。因为可能出现多个字母频率相同的情况，在反转的字典中，每项的值应当是字母的列表。

这里是一个反转字典的函数：

```
def invert_dict(d):
    inverse = dict()
    for key in d:
        val = d[key]
        if val not in inverse:
            inverse[val] = [key]
        else:
            inverse[val].append(key)
    return inverse
```

每次循环中，`key`从`d`中获得一个键，而`val`获得相应的值。如果`val`不在`inverse`字典中，意味着我们还没有见到过它，所以新建一个项，并将它初始化为一个单件（singleton，即只包含一个元素的列表）。否则我们已经见过这个值了，因此将相应的键附加到列表末尾。

下面是一个示例：

```
>>> hist = histogram('parrot')
>>> print hist
{'a': 1, 'p': 1, 'r': 2, 't': 1, 'o': 1}
>>> inverse = invert_dict(hist)
>>> print inverse
{1: ['a', 'p', 't', 'o'], 2: ['r']}
```

图11-1是一个显示`hist`和`inverse`的状态图。字典使用一个上方标明`dict`的图框表示，内部包含键值对。如果值是整数、浮点数或字符串，我通常会把它们画到图框内，但我常常会将列表画在图框之外，以便保持状态图的简洁。

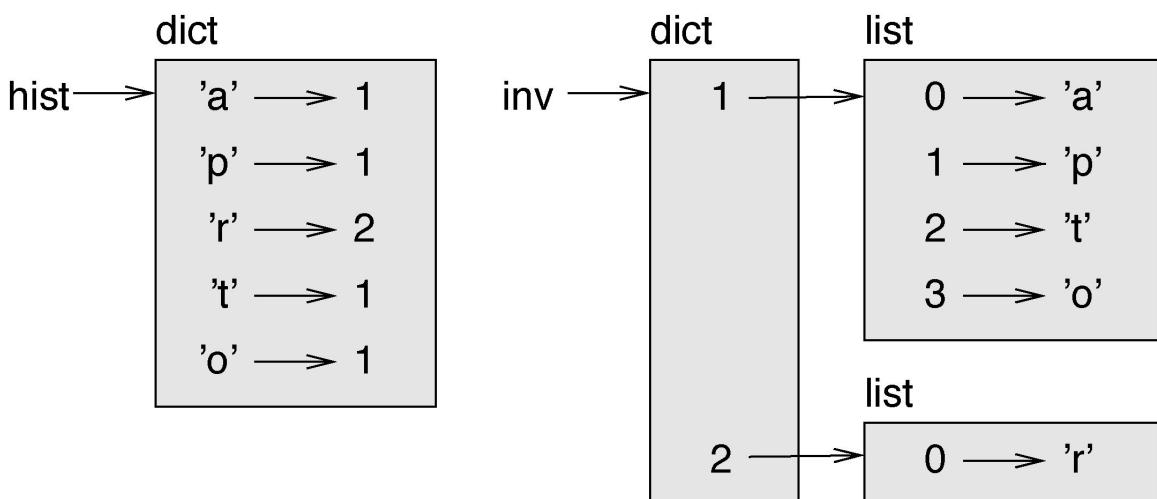


图11-1 状态图

如本例所示，列表可以用作字典的值，但它们不能用作键。如果你尝试的话，会得到如下的结果：

```
>>> t = [1, 2, 3]
>>> d = dict()
>>> d[t] = 'oops'
Traceback (most recent call last):
  File "<stdin>", line 1, in ?
TypeError: list objects are unhashable
```

之前我提到过字典是通过散列表的方式实现的，这意味着键必须是可散列（hashable）的。

散列是一个函数，接收（任意类型）的值并返回一个整数。字典使用这些被称为散列值的整数来保存和查找键值对。

这套系统当键不可变时，可以正确工作。但如果像列表这样，键是可变的话，则会有不好的事情发生。比如，当你新建一个键值对，Python将键进行散列并存储到对应的地方。如果你修改了键并再次散列，它会指向一个不同的地方。在那种情况下，会导致同一个键有两个条目，或者可能找不到某个键。不论如何，字典将无法正确工作。

因此键必须是可散列的，而类似列表这样的可变类型是不可散列的。绕过这种限制的最简单办法是使用元组，下一章会有详细介绍。

因为字典是可变的，它也不能用作键，但它可以用作字典的值。

练习11-5

阅读字典方法setdefault的文档，并使用它来写一个更简洁的 invert\_dict。解答：[http://thinkpython.com/code/invert\\_dict.py](http://thinkpython.com/code/invert_dict.py)。

## 11.5 备忘

如果你尝试过6.7节中的fibonacci函数，可能会注意到，提供的参数越大，函数运行的时间越长，并且运行时间增长地非常快。

为了明白为什么会这样，考虑图11-2，它展示了fibonocci函数n=4时的调用图。

调用图显示了一组函数帧，并用箭头将函数的帧和它调用的函数帧连接起来。在图的顶端，n=4的fibonacci调用了n=3和n=2的fibonacci。同样地，n=3的fibonacci调用了n=2和n=1的fibonacci。依此类推。

数一下fibonacci(0)和fibonacci(1)被调用了多少次。这是本问题的一个很无效的解决方案，而且当参数变大时，事情会变得更糟。

一个解决办法是记录已经计算过的值，并将它们保存在一个字典中。将之前计算的值保存起来以便后面使用的方法称为备忘（memo）。下面是一个使用了备忘的fibonacci实现：

```
known = {0:0, 1:1}

def fibonacci(n):
    if n in known:
        return known[n]

    res = fibonacci(n-1) + fibonacci(n-2)
    known[n] = res
    return res
```

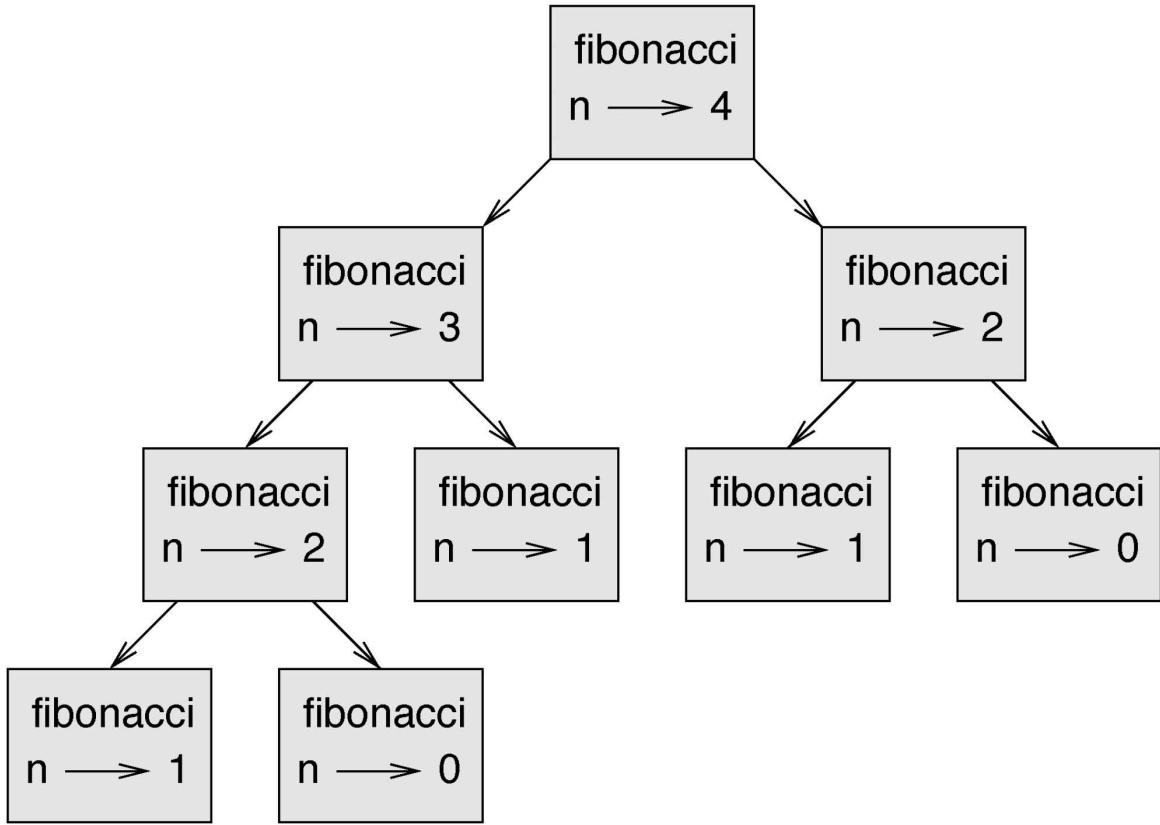


图11-2 调用图

`known`是一个用来记录我们已知的Fibonacci数的字典。开始时它有两项：0映射到0，以及1映射到1。

每当`fibonacci`被调用时，它会先检查`known`。如果结果已经存在，则可以立即返回。如果不存在，它需要计算这个新值，将其添加进字典，并返回。

#### 练习11-6

使用不同的实参运行这个版本和原版的`fibonacci`函数，并对比它们的运行时间。

#### 练习11-7

将练习6-5中的Ackermann函数改为使用备忘的版本，并查看备忘化之后是否能让它运行更大的参数。提示：no。解答：

[http://thinkpython.com/code/ackermann\\_memo.py](http://thinkpython.com/code/ackermann_memo.py)。

## 11.6 全局变量

在前一个例子中，`known`是在函数之外创建的，所以它属于被称为`__main__`的特殊帧。`__main__`之中的变量有时被称为全局变量，因为它可以在任意函数中访问。和局部变量在函数结束时就消失不同，全局变量可以在不同函数的调用之间持久存在。

全局变量常常用作标志（flags）；它是一种布尔变量，可以标志一个条件是否为真。比如，有的函数使用一个叫`verbose`的标志来控制输出的详细程度：

```
verbose = True

def example1():
    if verbose:
        print 'Running example1'
```

如果你尝试给全局变量重新赋值，可能会感到惊讶。下面例子的本意是想记录函数是否被调用过：

```
been_called = False

def example2():
    been_called = True      # 错
```

但当你运行它时，会发现`been_called`的值并不会变化。问题在于函数`example2`会新建一个局部变量`been_called`。局部变量在函数结束时就会消失，并且对全局变量没有任何影响。

要想在函数中给全局变量重新赋值，你需要在使用它之前先声明这个全局变量：

```
been_called = False

def example2():
    global been_called
    been_called = True
```

`global`语句告诉编译器，“在这个函数里，当我说`been_called`时，我指的是全局变量；不要新建一个局部变量。”

下面是一个尝试更新全局变量的例子：

```
count = 0

def example3():
    count = count + 1          # 错
```

如果运行它，会得到：

```
UnboundLocalError: local variable 'count' referenced before assignment
```

Python会认为count是局部的，这样意味着你在写入它之前先读取了。解决方案也是声明count为全局变量。

```
def example3():
    global count
    count += 1
```

如果全局变量是可变的，你可以不用声明它就进行修改：

```
known = {0:0, 1:1}

def example4():
    known[2] = 1
```

所以你可以添加、删除和替换一个全局的列表或字典的元素，但如果想要给全局变量重新赋值，则需要声明它：

```
def example5():
    global known
    known = dict()
```

## 11.7 长整数

如果计算fibonacci(50)，会得到：

```
>>> fibonacci(50)
12586269025L
```

数字结尾的L表明结果是一个长整数，或者说类型是long。在Python 3中，long已经消失了；所有的整数，包括特别大的，都是类型int。

类型为int的值有有限的范围；长整数则可以无限大，但当它们变大时，会占用更多的空间和时间。

数学操作符可以用于长整数，并且math模块中的所有函数也可以。因此大体来说，任何能够处理int的函数都可以处理long。

任何时候，如果计算的结果太大，以至于无法用int整数表示，Python会将它转换为一个长整数：

```
>>> 1000 * 1000
1000000
>>> 100000 * 100000
10000000000L
```

在第一个例子里结果是int类型；第二个的结果是long。

### 练习11-8

大整数的乘方是常用的公钥加密算法的基础。阅读RSA算法的维基百科（<http://en.wikipedia.org/wiki/RSA>）并编写函数来加密和解密信息。

## 11.8 调试

当你使用更大的数据集时，通过打印和手动检查数据的方式来调试已经变得很笨拙了。下面是一些调试大数据集的建议。

- **缩小输入：**如果可能，减小数据集的尺寸。比如，程序如果读入文本文件，可以从开头10行开始，或者使用你能找到的最小样本。你可以编辑文件本身，或者（更好地）修改程序让它只读入前 $n$ 行。  
如果出现了错误，你可以调小 $n$ ，小到足够展现出错误的最小程度，并在修改之后逐渐增大 $n$ 。
- **检查概要信息和类型：**与其打印和检查整个数据集，可以考虑打印出数据的概要信息：例如，字典中条目的数量，或者一个列表中数的和。  
运行时错误的一个常见原因是某个值的类型不对。调试这种错误时，常常只需要打印出值的类型就足够了。
- **编写自检查逻辑：**有时候可以写代码自动检查错误。例如，如果你要计算一系列数的平均值，可以检查结果是否比列表中最大的数小，或者比最小的数大。这种检查称为“健全检查”（sanity check），因为它会发现那些“发疯”的结果。  
另一种检查可以对比两种不同的计算的结果，查看它们是否一致。这样的检查称为“一致性检查”。
- **美化输出：**格式化调试输出，可以更容易发现错误。我们在6.9节中已经看到过一个例子。`pprint`模块提供了一个`pprint`函数，可以将内置类型的值以更加人性化的可读的格式打印出来。  
另外，再提醒一次，花费时间构建脚手架代码，可以减少未来进行调试的时间。

## 11.9 术语表

字典（dictionary）：从键的集合到对应的值的集合的映射。

键值对（key-value pair）：键到值的映射的展示。

项（item）：键值对的另一个名称。

键（key）：字典中出现在键值对的前一部分的对象。

值（value）：字典中出现在键值对的后一部分的对象。这比我们之前提到的“值”更加具体。

实现（implementation）：进行计算的一个具体方式。

散列表（hashtable）：Python字典的实现用的算法。

散列函数（hash function）：散列表中用来计算一个键的位置的函数。

可散列（hashable）：拥有散列函数的类型。不可变类型，诸如整数、浮点数和字符串都是可散列的；可变类型，诸如列表和字典，都是不可散列的。

查找（lookup）：字典的一个操作，接收一个键，并找到它对应的值。

反向查找（reverse lookup）：字典的一个操作，通过一个值来找到它对应的一个或多个键。

单件（singleton）：只包含一个元素的列表（或其他序列）。

调用图（call graph）：一个用来展示程序运行中创建的每一帧的关系的图。使用箭头连接每个调用者和被调用者。

直方图（histogram）：一个计数器的集合。

备忘（memo）：将计算的结果存储起来，以避免将来进行不必要的计算。

全局变量（global variable）：在函数之外定义的变量。全局变量可以在任何函数中访问。

标志（flag）：用于标志一个条件是否为真的布尔变量。

声明（declaration）：类似于global这样的用于通知解释器关于一个变量的信息的语句。

## 11.10 练习

### 练习11-9

如果你做过练习10-8，则已经有一个接受了列表作为形参的函数 `has_duplicates`，当列表中有任意元素出现多于1次时返回True。

使用字典编写一个更快、更简单的`has_duplicates`。

解答：[http://thinkpython.com/code/has\\_duplicates.py](http://thinkpython.com/code/has_duplicates.py)。

### 练习11-10

两个单词，如果可以使用轮转操作将一个转换为另一个，则称为“轮转对”（参见练习8-12中的`rotate_word`函数）。

编写一个程序，读入一个单词表，并找到所有的轮转对。

解答：[http://thinkpython.com/code/rotate\\_pairs.py](http://thinkpython.com/code/rotate_pairs.py)。

### 练习11-11

下面是《车迷天下》节目中的另一个谜题

(<http://www.cartalk.com/content/puzzler/transcripts/200717>)：

这个谜题是一个叫做Dan O'Leary的伙计寄过来的。他曾经遇到一个单音节、5字母的常用单词，有如下所述的特殊属性。当你删除第一个字母时，剩下的字母组成原单词的一个同音词，即发音完全相同的词。将第一个字母放回去，并删除第二个字母，结果也是原单词另一个同音词。问题是，这个单词是什么？

接下来我给你一个示例，但它并不能完全符合条件。我们看这个5字母单词“wrack”，W-R-A-C-K，也就是“wrack with pain”（“带来伤害”）里的那个词。如果我删掉第一个字母，会剩下四个字母的单词，“R-A-C-K”。也就是，“Holy cow, did you see the rack on that buck! It must have a nine-pointer!”（“天哪！你看到那匹雄鹿的鹿角了吗！一定有9个犄角！”）里的那个词。它是一个完美的同音词。但如果你把“w”放回去，并删掉“r”，会得到单词“wack”，也是一个真实单词，但它读音和其他两个不一样。

但就Dan和我所知，至少有一个单词能够通过删除前两个字母得到两个同音词。问题是，这个单词是什么？

你可以使用练习11-1中的字典来检测一个字符串是否出现在单词表中。

要检查两个单词是不是同音词，可以使用CMU发音词典。你可以从这里 (<http://www.speech.cs.cmu.edu/cgi-bin/cmudict>) 或者这里 (<http://thinkpython.com/code/c06d>) 下载它，也可以下载 <http://thinkpython.com/code/pronounce.py>，其中提供了一个叫做 `read_dictionary` 的函数来读入发音词典并返回一个Python字典，将每个单词映射到表示其主要读音的字符串上。

编写一个程序，列出所有可以解答这个谜题的单词。

解答：<http://thinkpython.com/code/homophone.py>。

# 第12章 元组

## 12.1 元组是不可变的

元组是值的一个序列。其中的值可以是任何类型，并且按照整数下标索引，所以从这方面看，元组和列表很像。元组和列表之间的重要区别是，元组是不可变的。

语法上，元组就是用逗号分隔的一列值：

```
>>> t = 'a', 'b', 'c', 'd', 'e'
```

虽然并不必需，但元组常常用括号括起来：

```
>>> t = ('a', 'b', 'c', 'd', 'e')
```

若要新建只包含一个元素的元组，需要在后面添加一个逗号：

```
>>> t1 = 'a',
>>> type(t1)
<type 'tuple'>
```

而用括号括起来的单独的值并不是元组：

```
>>> t2 = ('a')
>>> type(t2)
<type 'str'>
```

新建元组的另一种形式是使用内置函数tuple。不带参数时，它会新建一个空元组：

```
>>> t = tuple()
>>> print t
()
```

如果参数是一个序列（字符串、列表或者元组），结果就是一个包含序列的元素的元组：

```
>>> t = tuple('lupins')
>>> print t
('l', 'u', 'p', 'i', 'n', 's')
```

因为tuple是内置函数的名称，所以应当避免用它作为变量名称。

大多数列表操作也可以用于元组。方括号操作符可以用下标取得元素：

```
>>> t = ('a', 'b', 'c', 'd', 'e')
>>> print t[0]
'a'
```

而切片操作符选择一个范围内的元素：

```
>>> print t[1:3]
('b', 'c')
```

但如果尝试修改元组中的一个元素，会得到错误：

```
>>> t[0] = 'A'
TypeError: object doesn't support item assignment
```

不能修改元组的元素，但是可以将一个元组替换为另一个：

```
>>> t = ('A',) + t[1:]
>>> print t
('A', 'b', 'c', 'd', 'e')
```

## 12.2 元组赋值

交换两个变量的值常常很有用。使用传统的赋值方式，需要使用一个临时变量。例如，要交换a和b：

```
>>> temp = a  
>>> a = b  
>>> b = temp
```

这种解决方案很笨拙，而元组赋值则更优雅：

```
>>> a, b = b, a
```

左边是一个变量的元组，右边是表达式的元组。每个值会被赋值给相应的变量。右边所有的表达式，都会在任何赋值操作进行之前完成求值。

左边变量的个数和右边值的个数必须相同：

```
>>> a, b = 1, 2, 3  
ValueError: too many values to unpack
```

更通用地，右边可以是任意类型的序列（字符串、列表或元组）。例如，想要将电子邮件地址拆分成用户名和域名，可以这么写：

```
>>> addr = 'monty@python.org'  
>>> uname, domain = addr.split('@')
```

split返回两个元素的列表；第一个元素被赋值到uname，第二个到domain上。

```
>>> print uname  
monty  
>>> print domain  
python.org
```

## 12.3 作为返回值的元组

严格地说，函数只能返回一个值，但如果返回值是元组的话，效果和返回多个值差不多。例如，如果将两个整数相除，得到商和余数，那么先计算 $x/y$ 再计算 $x \% y$ 并不高效。更好的方法是同时计算它们。

内置函数`divmod`接收两个参数，并返回两个值的元组，即商和余数。可以将结果存为一个元组：

```
>>> t = divmod(7, 3)
>>> print t
(2, 1)
```

或者可以使用元组赋值来分别存储结果中的元素：

```
>>> quot, rem = divmod(7, 3)
>>> print quot
2
>>> print rem
1
```

下面是返回一个元组的函数的示例：

```
def min_max(t):
    return min(t), max(t)
```

`max`和`min`都是内置函数，分别返回一个序列的最大值和最小值。`min_max`计算这两个值并将它们作为一个元组返回。

## 12.4 可变长参数元组

函数可以接收不定个数的参数。以`*`开头的参数名会收集（gather）所有的参数到一个元组上。例如，`printall`接收任意个数的参数并打印它们：

```
def printall(*args):
    print args
```

收集参数可以使用任何你想要的名称，但按惯例通常使用`args`。下面是函数如何工作的一个例子：

```
>>> printall(1, 2.0, '3')
(1, 2.0, '3')
```

收集的反面是分散（scatter）。如果有一个序列的值而想将它们作为可变长参数传入到函数中，可以使用`*`操作符。例如，`divmod`正好接收两个参数，但它不接收元组：

```
>>> t = (7, 3)
>>> divmod(t)
TypeError: divmod expected 2 arguments, got 1
```

但如果将元组分散，就可以用了：

```
>>> divmod(*t)
(2, 1)
```

### 练习12-1

很多内置函数使用可变长参数元组。例如，`max`和`min`都可以接收任意个数的参数：

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

但是`sum`并不这样。

```
>>> sum(1, 2, 3)
TypeError: sum expected at most 2 arguments, got 3
```

编写一个函数sumall，接收任意个数的参数并返回它们的和。

## 12.5 列表和元组

`zip`是一个内置函数，接收两个或多个序列，并将它们“拉”到一起，成为一个元组列表。每个元组包含各个序列中的一个元素。在Python 3中，`zip`返回元组的迭代器，不过大多数情况下，迭代器的行为和列表差不多。

下面的例子将字符串和一个列表“拉”到一起：

```
>>> s = 'abc'  
>>> t = [0, 1, 2]  
>>> zip(s, t)  
[('a', 0), ('b', 1), ('c', 2)]
```

结果是元组的列表，每个元组包含字符串中的一个字符和列表中对应的元素。

如果序列之间的长度不同，则结果的长度是所有序列中最短的那个。

```
>>> zip('Anne', 'Elk')  
[('A', 'E'), ('n', 'l'), ('n', 'k')]
```

可以在`for`循环中使用元组赋值来访问元组的列表：

```
t = [('a', 0), ('b', 1), ('c', 2)]  
for letter, number in t:  
    print number, letter
```

每次循环中，Python选择列表中的下一个元组，并将其元素赋值给`letter`和`number`变量。这个循环的输出如下：

```
0 a  
1 b  
2 c
```

如果组合使用`zip`、`for`循环以及元组赋值，可以得到一种有用的模式，用于同时遍历两个或更多序列。例如，`has_match`函数接收两个序列，`t1`和`t2`，并当存在一个下标`i`保证`t1[i] == t2[i]`时返回`True`：

```
def has_match(t1, t2):  
    for x, y in zip(t1, t2):
```

```
if x == y:  
    return True  
return False
```

如果需要遍历序列中的元素以及它们的下标，可以使用内置函数 `enumerate`:

```
for index, element in enumerate('abc'):  
    print index, element
```

同样，这个循环的输出是:

```
0 a  
1 b  
2 c
```

## 12.6 字典和元组

字典有一个`items`方法可以返回一个元组的列表，其中每个元组是一个键值对。

```
>>> d = {'a':0, 'b':1, 'c':2}
>>> t = d.items()
>>> print t
[('a', 0), ('c', 2), ('b', 1)]
```

和预料中一样，字典中的项是没有特定顺序的。在Python 3中，`items`返回一个迭代器，但在很多场景中，迭代器和列表差不多。

从反方向出发，可以使用一个元组列表来初始化一个新的字典：

```
>>> t = [('a', 0), ('c', 2), ('b', 1)]
>>> d = dict(t)
>>> print d
{'a': 0, 'c': 2, 'b': 1}
```

组合使用`dict`和`zip`可以得到一个简洁的创建字典的方法：

```
>>> d = dict(zip('abc', range(3)))
>>> print d
{'a': 0, 'c': 2, 'b': 1}
```

字典方法`update`也接收一个元组列表，并将它们作为键值对添加到一个已有的字典中。

组合使用`items`、元组赋值和`for`循环，可以直接遍历字典中的键和值：

```
for key, val in d.items():
    print val, key
```

这个循环的输出也是：

```
0 a
2 c
1 b
```

使用元组作为字典的键很常见（主要是因为不能使用列表）。例如，一个电话号码簿可能需要将姓名对映射到电话号码。假设定义了`last`, `first`和`number`, 可以这么写：

```
directory[last,first] = number
```

在方括号中的表达式是一个元组。我们也可以使用元组赋值来遍历这个字典：

```
for last, first in directory:  
    print first, last, directory[last,first]
```

这个循环遍历字典`directory`的所有键，它们都是元组。它将每一个元组的元素赋值给`last`和`first`, 接着打印出名字以及对应的电话号码。

在状态图中有两种方法可以表达元组。更详细的版本和列表一样，显示索引和元素。例如，元组`('Cleese', 'John')`可以如图12-1所示。

## tuple

```
0 → 'Cleese'  
1 → 'John'
```

图12-1 状态图

但是在更大的图中你可能希望省略掉细节。例如，整个电话簿的图如图12-2所示。

dict

```
('Cleese', 'John') —> '08700 100 222'  
('Chapman', 'Graham') —> '08700 100 222'  
    ('Idle', 'Eric') —> '08700 100 222'  
    ('Gilliam', 'Terry') —> '08700 100 222'  
    ('Jones', 'Terry') —> '08700 100 222'  
    ('Palin', 'Michael') —> '08700 100 222'
```

图12-2 状态图

这里元组使用Python的语法作为图形化的简写展示。

这张图里的电话号码是BBC的投诉热线，所以请不要真去拨打它。

## 12.7 比较元组

关系操作符对元组和其他序列都可用；Python会从每一个序列的第一个元素开始比较。如果它们相等，则继续比较下一个元素，如此类推，直到找到不同的元素。后面的元素就不考虑了（即使它们非常大）。

```
>>> (0, 1, 2) < (0, 3, 4)
True
>>> (0, 1, 2000000) < (0, 3, 4)
True
```

`sort`函数也用相同的方式工作。它主要按照第一个元素排序，但如果两者相同，则会按第二个元素排序，如此类推。

这个特点形成一个称为DSU的模式。

- 修饰（Decorate）：构建一个元组列表，在序列元素之前放置一个或多个排序键。
- 排序（Sort）：给这个序列排序，并。
- 去修饰（Undecorate）：抽取排好序的序列中的元素。

例如，假设有一个单词列表，想要按照它们的长度从大到小来排序：

```
def sort_by_length(words):
    t = []
    for word in words:
        t.append((len(word), word))

    t.sort(reverse=True)

    res = []
    for length, word in t:
        res.append(word)
    return res
```

第一个循环构建一个元组列表，每个元组是一个单词长度以及对应的单词。

`sort`会先比较第一个元素，也就是长度，并只有当相同时才会考虑第二个元素。键值参数`reverse=True`告诉`sort`按照降序来排序。

第二个循环遍历元组列表，并构建一个按照长度降序的单词列表。

## 练习12-2

在这个例子中，当遇到相等时，通过比较单词来确定顺序，所以长度相同的单词会按照字母顺序的反序排列。对于其他的程序，你可能想要随机处理相等的情况。修改这个示例，当单词长度相等时，单词随机排序。

提示：参考random模块的random函数。解答：

[http://thinkpython.com/code/unstable\\_sort.py](http://thinkpython.com/code/unstable_sort.py)。

## 12.8 序列的序列

我一直在聚焦于元组的列表，但本章中几乎所有的示例都可以对列表的列表、元组的元组以及列表的元组使用。为了避免枚举所有的可能组合，有时候直接说序列的序列更简单。

在很多环境中，不同类型的序列（字符串、列表和元组）都可以互换使用。应当如何选择使用哪个呢？

从最明显的一个开始，字符串比其他序列有更多限制，因为它的元素必须是字符。它们也是不可变的。如果你需要修改一个字符串中的字符（而不是新建一个字符串），可能需要使用字符的列表。

列表比元组更加通用，主要因为它是可变的。但也有一些情况下你可能会优先选择元组：

1. 在有些环境中，比如返回语句中，创建元组比创建列表从语法上说更容易。其他的环境中，你可能会选择列表。
2. 如果需要用序列作为字典的键，则必须使用不可变类型，比如元组或字符串。
3. 如果你要向函数传入一个序列作为参数，使用元组可能会减少潜在的由假名导致的不可预知行为。

因为元组是不可变的，它们不提供类似`sort`和`reverse`之类的方法，这些方法需要修改序列。但Python也提供了内置函数`sorted`和`reversed`，可以接收任何类型的序列作为参数，并返回一个排好序的新序列。

## 12.9 调试

列表、字典和元组都被统一看做是一种数据结构；本章中我们开始看到复合数据结构，像元组的列表，以及用元组做键、用列表做值的字典等。复合数据结构很有用，但它容易导致我称为的结构错误；也就是说，数据结构因为错的类型、大小或组合导致的错误。例如，如果你期望得到一个包含单个整数的列表，而我给你一个单个整数（而不是在列表中），就会出错。

为了帮助调试这种问题，我写了一个模块`structshape`，提供一个也叫做`structshape`的函数，接收任何数据类型作为参数，并返回一个描述它的形状的字符串。你可以从<http://thinkpython.com/code/structshape.py>下载它。

下面是一个简单列表的结果：

```
>>> from structshape import structshape
>>> t = [1, 2, 3]
>>> print structshape(t)
list of 3 int
```

更好看的程序可能会输出“list of 3 ints”，但不需要处理复数更加容易。下面是列表的列表：

```
>>> t2 = [[1, 2], [3, 4], [5, 6]]
>>> print structshape(t2)
list of 3 list of 2 int
```

如果列表的元素不是同一种类型，`structshape`会根据它们的类型按顺序分组：

```
>>> t3 = [1, 2, 3, 4.0, '5', '6', [7], [8], 9]
>>> print structshape(t3)
list of (3 int, float, 2 str, 2 list of int, int)
```

下面是元组的列表：

```
>>> s = 'abc'
>>> lt = zip(t, s)
>>> print structshape(lt)
list of 3 tuple of (int, str)
```

下面是一个字典，有3个从整数映射到字符串的项：

```
>>> d = dict(lt)
>>> print structshape(d)
dict of 3 int->str
```

如果你发现要记住数据结构有困难，`structshape`可以帮忙。

## 12.10 术语表

元组 (tuple) : 一个不可变的元素序列。

元组赋值 (tuple assignment) : 一个赋值语句，右侧是一个序列，左侧是一个变量的元组。右边的序列会被求值，它的元素依次赋值给左侧元组中的变量。

收集 (gather) : 组装可变长度的参数元组的操作。

分散 (scatter) : 把一个序列当做参数列表的操作。

DSU: “decorate-sort-undecorate”的缩写，构造一个元组列表，排序，再抽取部分结果的模式。

数据结构 (data structure) : 相关的值的集合，通常组织成列表、字典、元组等。

数据结构的构形 (shape of a data structure) : 数据结构的类型、尺寸和组合的摘要。

## 12.11 练习

### 练习12-3

编写一个函数`most_frequent`，接收一个字符串并按照频率的降序打印字母。从不同语言中查找文本样例并查看不同语言中的单词频率如何变化。将你的结果和[http://en.wikipedia.org/wiki/Letter\\_frequencies](http://en.wikipedia.org/wiki/Letter_frequencies)上的列表进行对比。

解答：[http://thinkpython.com/code/most\\_frequent.py](http://thinkpython.com/code/most_frequent.py)。

### 练习12-4

更多回文！

1. 编写一个程序从文件中读入一个单词列表（参见9.1节）并打印出所有是回文的单词集合。

下面是输出的样子的示例：

```
['deltas', 'desalt', 'lasted', 'salted', 'slated', 'staled']
['retainers', 'ternaries']
['generating', 'greatening']
['resmelts', 'smelters', 'termless']
```

提示：

你可能需要构建一个字典将字母的集合映射到可以用这些字母构成的单词的列表上。问题是，如何表达字母集合，才能让它可以用作字典的键？

2. 修改前一个问题的程序，让它先打印最大的回文集合，再打印第二大的回文集合，依次类推。

3. 在Scrabble拼字游戏中，一个“bingo”代表你自己架子上全部7个字母和盘上的一个字母组合成一个8字母单词。哪一个8字母单词可以生成最多的bingo？提示：一共有7个。

解答：[http://thinkpython.com/code/anagram\\_sets.py](http://thinkpython.com/code/anagram_sets.py)。

### 练习12-5

两个单词，如果可以通过交换两个字母将一个单词转换为另一个，就称为“置换对”；例如，“converse”和“conserve”。编写一个程序查找字典中所有的置换对。提示：不要测试所有的单词对，也不要测试所有可能的交换。

解答：<http://thinkpython.com/code/metathesis.py>。

鸣谢：这个练习启发自<http://puzzlers.org>的示例。

### 练习12-6

下面是《车迷天下》节目中的一个谜题  
(<http://www.cartalk.com/content/puzzler/transcripts/200651>)：

一个英文单词，当逐个删除它的字母时，仍然是英文单词。这样的单词中最长的是什么？

首先，字母可以从两头或者中间删除，但你不能重排字母。每次你去掉一个字母，则得到另一个英文单词。如果一直这么做，最终会得到一个字母，它本身也是一个英文单词——可以从字典上找到的。我想知道这样的最长的单词是什么，它有多少字母？

我会给你一个普通的例子：Sprite。你从sprite开始，取出一个字母，从单词内部取，取走r，这样我们就剩下单词spite，接着我们取走结尾的e，剩下spit，接着取走s，我们剩下pit、it和I。

编写一个程序来找到所有可以这样缩减的单词，然后找到最长的一个。

这个练习比大部分练习都更有挑战，所以下面有一些建议。

1. 你可能需要编写一个程序接收一个单词，并计算出所有通过从它取出一个字母得到的单词的列表。它们是这个单词的“子”单词。
2. 递归地，只有当一个单词的子单词中有一个可缩减时，它才可缩减。作为一个基准情形，你可以认为空字符串可缩减。
3. 我提供的单词表，`words.txt`，并不存在单个字母的单词。所以你可能需要加上“I”、“a”和空字符串。
4. 为了提高程序的效率，你可能需要记住已知的可缩减的单词。

解答：<http://thinkpython.com/code/reducible.py>。

# 第13章 案例研究：选择数据结构

## 13.1 单词频率分析

和前面的章节一样，你应当至少尝试一下解决下面的问题，再看我的解答。

### 练习13-1

编写一个程序，读入一个文件，将每行内容拆解为单词，剥去单词周围的空白字符和标点，并转换为小写。

**提示：**

`string`模块提供了空白字符串`whitespace`，包括空格、制表符、换行符等；它也提供了`punctuation`，包含了所有的标点字符。让我们试试能不能让Python胡言乱语：

```
>> import string  
>> print string.punctuation  
!"#$%&'()*+,-./:;=>?@[\\]^_`{|}~
```

另外，你也可以考虑字符串方法`strip`、`replace`和`translate`。

### 练习13-2

去往古腾堡工程（Project Gutenberg，<http://www.gutenberg.org>）并下载你最喜欢的无版权书籍的纯文本文档。

修改前一个练习中的程序，改为从你下载的书籍中读取内容，跳过文件开头的信息部分，并和前面一样将文本处理成为单词。

接着修改程序，计算书中出现的全部单词的总数，以及每个单词使用的次数。

打印书中使用的不同单词的个数。比较不同时代、不同作者的不同书籍。哪一个作者使用的词汇最广泛？

#### 练习13-3

修改前一个练习中的程序，计算书中使用频率最高的20个单词。

#### 练习13-4

修改前面的程序，读入一个单词表（参见9.1节）并打印出书中所有不在单词表之中的单词。这其中有多少是拼写错误？有多少是应该出现在单词表中的常用单词？有多少是真正冷僻的单词？

## 13.2 随机数

给定相同的输入，大部分计算机程序每次运行都会生成相同的输出，所以它们被认为是有确定性的。确定性通常是件好事，因为我们希望相同的计算能有相同的结果。但对某些特别的应用，我们希望计算机是不可预测的。游戏是一个明显的例子，但还有更多类似的例子。

让程序变得真正地不确定，其实并不是容易的事情。但也有办法让它最少看起来是不确定的。一种办法是使用算法来生成伪随机数。伪随机数并不是真正随机的，因为它们是通过一个确定性的算法生成的，但若只看输出的数字的话，几乎不可能看出来和随机数有什么区别。

模块random提供了用于生成伪随机数的函数（接下来我直接简单地将它称为“随机数”）。

函数random返回一个从0.0到1.0之间的随机浮点数（包括0.0，但不包括1.0）。每当调用random时，会得到一个很长的随机数序列中的下一个数。运行下面的循环，可以看到一个样本：

```
import random

for i in range(10):
    x = random.random()
    print x
```

函数randint接收参数low和high，并返回low和high之间（两者都包含）的一个整数。

```
>>> random.randint(5, 10)
5
>>> random.randint(5, 10)
9
```

要从序列中随机选择一个元素，可以使用choice：

```
>>> t = [1, 2, 3]
>>> random.choice(t)
2
>>> random.choice(t)
3
```

random模块还提供了可以从各种连续分布序列中生成随机数的函数。包括高斯分布、指数分布、gamma分布，以及其他几种。

### 练习13-5

编写一个函数choose\_from\_hist，接收一个11.1节所定义的直方图作为参数，并从这个直方图中，按照频率的大小，成比例地随机返回一个值。例如，对下面这个直方图：

```
>>> t = ['a', 'a', 'b']
>>> hist = histogram(t)
>>> print hist
{'a': 2, 'b': 1}
```

你的函数应该以 $2/3$ 的几率返回'a'，以 $1/3$ 的几率返回'b'。

## 13.3 单词直方图

在继续阅读之前你应当尝试前面的练习。你可以从[http://thinkpython.com/code/analyze\\_book.py](http://thinkpython.com/code/analyze_book.py)下载我的解答。你还需要<http://thinkpython.com/code/emma.txt>。

下面是一个读取文件并从文件中的单词构造直方图的例子：

```
import string

def process_file(filename):
    hist = dict()
    fp = open(filename)
    for line in fp:
        process_line(line, hist)
    return hist

def process_line(line, hist):
    line = line.replace('-', ' ')
    for word in line.split():
        word = word.strip(string.punctuation + string.whitespace)
        word = word.lower()
        hist[word] = hist.get(word, 0) + 1

hist = process_file('emma.txt')
```

这个程序读入`emma.txt`，其内容是简·奥斯丁的《爱玛》的文本。

`process_file`循环遍历文件中的每一行，每次将一行传递给`process_line`函数。直方图`hist`用作累加器。

`process_line`使用字符串方法`replace`将'-'符号替换为空格，再使用`split`将各行文本拆分成一个字符串列表。它遍历单词列表，使用`strip`和`lower`去除掉标点符号并转换为小写。（我们说“转换”，只是个简称，记住字符串是不可变的，所以`strip`和`lower`这样的方法返回的是新字符串。）

最后，`process_line`通过创建新项或者增加旧有项的值来更新直方图。

要计算文件中单词的总数，我们可以累加直方图中的频率：

```
def total_words(hist):  
    return sum(hist.values())
```

不同单词的个数，就是字典里的元素数量：

```
def different_words(hist):  
    return len(hist)
```

下面是打印结果的代码：

```
print 'Total number of words:', total_words(hist)  
print 'Number of different words:', different_words(hist)
```

以及结果：

```
Total number of words: 161080  
Number of different words: 7214
```

## 13.4 最常用的单词

要寻找最常用单词，我们可以应用DSU模式；`most_common`接收一个直方图，并返回“单词-频率”元组的列表，按频率高低反向排序：

```
def most_common(hist):
    t = []
    for key, value in hist.items():
        t.append((value, key))

    t.sort(reverse=True)
    return t
```

下面的循环打印出最常用的10个单词：

```
t = most_common(hist)
print 'The most common words are:'
for freq, word in t[0:10]:
    print word, '\t', freq
```

以及《爱玛》的结果：

```
The most common words are:
to 5242
the 5205
and 4897
of 4295
i 3191
a 3130
it 2529
her 2483
was 2400
she 2364
```

## 13.5 可选形参

我们已经见过一些接收可变数量的形参的内置函数和方法。用户也可以编写接收可选形参的自定义函数。例如，下面的函数打印一个直方图中最常见的单词：

```
def print_most_common(hist, num=10):
    t = most_common(hist)
    print 'The most common words are:'
    for freq, word in t[:num]:
        print word, '\t', freq
```

第一个形参是必需的；第二个是可选的。形参num的默认值是10。

如果你只提供一个实参：

```
print_most_common(hist)
```

num会获得默认值。如果你提供两个实参：

```
print_most_common(hist, 20)
```

num则会获得你所提供的实参值。换句话说，可选实参值覆盖默认形参值。

如果一个函数既有必需形参，也有可选形参，则所有的必需形参都必须在前面，后面跟着可选形参。

## 13.6 字典减法

寻找在书中出现却不在words.txt单词表中的单词，这个问题可以看做是集合减法；也就是说，我们想要找到出现在一个集合（书中的单词）而不在另一个集合（单词表中的单词）的所有单词。

subtract函数接收两个字典d1和d2，并返回一个新的字典，包含所有出现在d1中且不出现在d2中的键值。由于我们并不真的关心字典的值，我们将所有值都设为None。

```
def subtract(d1, d2):
    res = dict()
    for key in d1:
        if key not in d2:
            res[key] = None
    return res
```

要找出书中出现而在words.txt单词表中的词，我们可以使用process\_file为words.txt建立一个直方图，再使用减法：

```
words = process_file('words.txt')
diff = subtract(hist, words)

print "The words in the book that aren't in the word list are:"
for word in diff.keys():
    print word,
```

下面是《爱玛》一书中的部分结果：

```
The words in the book that aren't in the word list are:
rencontre jane's blanche woodhouses disingenuousness
friend's venice apartment ...
```

这些词中有些是名字或所有格单词。另外的，比如“rencontre”，已经不再常用。但也有一些是真应该包含在单词表中的！

练习13-6

Python提供了一个数据结构set，它提供了很多常见的集合操作。阅读<http://docs.python.org/lib/types-set.html>的文档，并编写一个程序使用集合减法来寻找出现在书中但不出现在单词表中的单词。解答：[http://thinkpython.com/code/analyze\\_book2.py](http://thinkpython.com/code/analyze_book2.py)。

## 13.7 随机单词

若要从直方图中随机选择一个单词，最简单的算法是根据计算得到的频率构建一个列表，其中每个单词根据词频有多个拷贝，并从中随机选择一个单词：

```
def random_word(h):
    t = []
    for word, freq in h.items():
        t.extend([word] * freq)

    return random.choice(t)
```

表达式`[word] * freq`创建一个列表，里面有单词`word`的`freq`个拷贝。`extend`方法和`append`类似，区别是接收的参数是一个序列。

### 练习13-7

这个算法可以使用，但效率并不高；每当选择一个随机单词时，它会重建列表，而这个列表和原书差不多长。一个明显的改进方法是只建立列表一次，再使用多次选择，但这么做列表仍然很大。

更好的替代方案如下。

1. 使用`keys`来获得书中所有的单词的列表。
2. 构建一个列表，包含单词频率的累积和（参见练习10-3）。这个列表中的最后一项是书中单词的总数`n`。
3. 在1到`n`之间随机选择一个数。使用二分查找（参见练习10-11）来找到随机数在累积和列表中应该出现的位置的下标。
4. 使用这个下标，在单词表中找到相应的单词。

编写一个程序，使用这个算法来从书中选择一个随机的单词。

解答：[http://thinkpython.com/code/analyze\\_book3.py](http://thinkpython.com/code/analyze_book3.py)。

## 13.8 马尔可夫分析

如果你从书中随机地获取单词，可以借此感受一下书中的词汇，但可能无法通过随机获取来得到一句话：

```
this the small regard harriet which knightley's it most things
```

一个随机单词的序列，很难组成有意义的话，因为相邻的词之间没有任何关联。例如，在一个真实的句子中，冠词“the”应当会后接一个形容词或名词，而不应是动词或副词。

测量这种类型的关联的方法之一是使用马尔可夫分析，它能够用于描述给定的单词的序列中下一个出现的单词的概率。例如，歌曲《Eric, the Half a Bee》的开头是：

Half a bee, philosophically, Must, ipso facto, half not  
be. But half the bee has got to be Vis a vis, its entity.  
D' you see? But can a bee be said to be Or not to be an  
entire bee When half the bee is not a bee Due to some ancient  
injury?

在这段文本中，短语“half the”总是后接着单词“bee”，但短语“the bee”则可能后接“has”或“is”。

马尔可夫分析的结果是一个从每个前缀（比如“half the”和“the bee”）到其所有可能后缀（比如“has”和“is”）的映射。

给定这种映射后，你就可以用它来生成随机文本。从任意前缀开始，并从它的可能后缀中随机选择一个。接着，你可以将前缀的结尾和后缀组合起来，作为下一个前缀，并继续重复。

例如，如果你以前缀“Half a”开始，则接下来一个单词必定是“bee”，因为这个前缀在文本中只出现了一次。下一个前缀是“a bee”，所以下一个后缀可能是“philosophically”、“be”或者“due”。

在这个例子中前缀的长度总是2，但其实你可以使用任意前缀长度来进行马尔可夫分析。前缀的长度被称为这个分析的“度”(order)。

### 练习13-8

马尔可夫分析：

1. 编写一个程序从文件中读入文本，并进行马尔可夫分析。结果应该是一个字典，将前缀映射到可能后缀的集合。集合可以是列表、元组或者字典；由你来做出合适的选择。你可以使用前缀长度2来测试程序，但编写程序时应当考虑可以方便地改为其他前缀长度。

2. 在前面编写的程序中添加一个函数，基于马尔可夫分析的结果随机生成文本。下面是一个从《爱玛》中使用前缀长度2生成的例子：

He was very clever, be it sweetness or be angry, ashamed or only amused, at such a stroke. She had never thought of Hannah till you were never meant for me?" "I cannot make speeches, Emma:" he soon cut it all himself.

对这个例子，我留下了每个单词后面的标点。结果几乎是语法正确的，但也不完全对。语义上，它看起来很像是有意义的，但也不完全是。

当增加前缀长度时，结果会怎么样？随机生成的文本会不会看来更有意义？

3. 一旦你的程序可以正常运行后，可以考虑尝试一下混搭：如果对两本或更多的书进行分析，则生成的随机文本会以一种有趣的方式混合各书中的词汇和短语。

致谢：本案例分析基于Kernighan和Pike的*The Practice of Programming* (Addison-Wesley, 1999) 书中的一个示例。

你应当在继续阅读前尝试这个练习，接着可从  
<http://thinkpython.com/code/markov.py> 下载我的解答。你也需要  
<http://thinkpython.com/code/emma.txt>。

## 13.9 数据结构

使用马尔可夫分析生成随机文本很有趣，但这个练习还有一个要点：数据结构的选择。在前面的练习中，你需要选择：

- 如何表达前缀；
- 如何表达可能的后缀的集合；
- 如何表达每个前缀到可能后缀的集合的映射。

最后一个选择很简单，我们已知的唯一的映射类型就是字典，所以这也是最自然的选择。

对前缀来说，最明显的选择是字符串、字符串列表或者字符串元组。对后缀来说，一种选择是列表，另一种是直方图（字典）。

你会如何选择？第一步需要思考每种数据结构需要实现的操作。对前缀而言，我们需要能够从前方删除一个单词，并在后方添加一个单词。例如，如果当前的前缀是“Half a”，而下一个单词是“bee”，则需要能够构造下一个前缀，“a bee”。

你的第一个选择可能是列表，因为列表添加和删除元素都很方便。但我们也需要使用前缀作为字典的键，所以列表被排除掉。对元组而言，虽然你不能附加或删除，但可以使用加法操作符来构建一个新的元组：

```
def shift(prefix, word):  
    return prefix[1:] + (word,)
```

`shift`接收一个单词的元组、`prefix`，以及一个字符串`word`，并构建一个新的元组，包含`prefix`中除了第一个之外的元素，并把`word`添加在最后。

对后缀集合而言，我们需要进行的操作包括添加一个新的后缀（或者增加一个已有后缀的频率），以及随机选择一个后缀。

添加一个新后缀，使用列表实现或者直方图实现效率上相同。从一个列表中随机选择元素很简单；从直方图中随机选择则更难一些（参见练习13-7）。

到此为止我们一直在讨论实现的简易性，但选择数据结构时，还有其他需要考虑的因素。一个是运行时间。有时候，我们可以从理论上预期一种数据结构比另一种更快；例如，我提到过`in`操作符，当元素数量很大时，在字典中使用比在列表中快。

但哪种实现会更快常常无法事先预知。一个办法是两种都实现，再比较哪个更快。这种方法称为基准比较（benchmarking）。比较实际的方案是先选择最容易实现的数据结构，然后看它是否对预期的程序而言足够快。如果已经足够，则不需要变动；否则，可以使用`profile`模块之类的工具，发现程序中哪些地方占用了最长的时间。

另一个考虑因素是存储空间。例如，使用直方图来保存后缀集合可能占用较少空间，因为不论一个单词在文本中出现多少次，你只需要保存一次。有的情况下，节省空间也可以让你的程序运行更快，而在极端的情形中，如果导致内存溢出，则程序无法正常运行。但对大多数程序来说，存储空间是次于运行速度的第二考虑因素。

最后一点：在这个讨论中，对于分析和生成两个过程，我暗示了我们应当使用相同的数据结构。但因为这是两个分开的阶段，所以也可以在分析阶段使用一种数据结构，再转换为另一种数据结构用于生成阶段。如果新的数据结构在生成阶段节省的时间大于转换花费的时间，则总的来说是有利的。

## 13.10 调试

当你在调试程序时，尤其是对付一个困难的bug时，可以尝试下面4点。

- 阅读：审阅你的代码，对自己读出来，并检查它是否和你想说的一致。
- 运行：做一些小修改并进行试验，或者运行不同的版本。通常如果在程序中正确的地方加上正确的输出，问题就会变得更加显而易见。但有时候你需要花费一些时间来做准备。
- 沉思：花些时间思考！可能是哪种类型的错误：语法的、运行时的、还是语义的？从错误消息或程序输出中可以得到什么信息？哪种错误可能导致你看到的问题？在问题出现之前，你最后的一次修改是什么？
- 回退：在某种情况下，最好的办法就是回退，取消掉最近的修改，直到你的程序恢复到之前没有错误且能够理解的程度。然后可以开始重新构建。

新手程序员有时会卡在这些环节中的某一个上，却忘了还可以尝试其他的环节。每个环节都有其独自的失败模式。

例如，当问题是一个拼写错误时，阅读代码可以帮忙，但若问题是概念误解导致，就没有效果了。如果你不理解你的程序，那么即使阅读100遍，也发现不了问题，因为错误是在你脑中的。

运行一些试验代码可以起到很大帮助，尤其是那些短小而简单的测试程序。但如果你没有思考或阅读代码就运行试验代码，则可能会陷入我称之为“随机走动编程”的模式之中。即毫无目标地随机改变程序，直到程序正确运行为止。毫无疑问，随机走动编程可能要花费很长的时间。

你需要花一定时间去思考。调试就像是一门实验科学。你应当至少有一个关于这个问题的假设。如果有两个以上的可能性，可以试着构思一个测试来排除其中一个。

暂停并休息一下也能够帮助思考。讨论也可以。如果你向其他人（或者甚至是你自己）解释这个问题，有时候会在描述问题结束之前就突然找到答案。

但如果有太多错误，或者你要修正的代码太太太复杂，即使最好的调试技巧也会失败。有时候最好的选择是回退，简化程序，直到得到一个你能够理解并且正确运行的程序。

新手程序员往往不愿意后撤，他们无法忍受删除一行代码（即使那是错误的代码）。如果能让你感觉更好，可以将程序复制到另外一个文件再来删减它。这样以后就可以一点一点地复制回来。

寻找一个困难的bug，需要阅读、运行、沉思，甚至有时候需要回退。如果你在这其中一个环节上卡住了，可以尝试其他的环节。

## 13.11 术语表

确定性 (deterministic) : 程序的一种特性：给定相同的输入，每次运行都会执行相同的操作。

伪随机 (pseudorandom) : 一序列数：看起来是随机的，但实际上是由带着确定性的程序生成的。

默认值 (default value) : 可选形参声明时给定的值，如果函数调用时没有指定这个实参的值，则使用该默认值。

覆盖 (override) : 使用实参值替换一个默认值。

基准测试 (benchmarking) : 实现不同的备选方案，并使用各种可能输入的样本来测试它们，以达到选择使用哪种数据结构的目的。

## 13.12 练习

### 练习13-9

一个单词的“排名”是它在单词列表中按频率排序的位置：最常见的词排名第1，次常用的词排第2，等等。

齐普夫定律 (Zipf's law) 描述了排名和自然语言中词频的关系 ([http://en.wikipedia.org/wiki/Zipf's\\_law](http://en.wikipedia.org/wiki/Zipf's_law))。特别地，它预测了排名为 $r$ 的单词的频率 $f$ :

$$f = cr^{-s}$$

这里 $s$  和 $c$  是依赖于语言和文本的参数。如果你在表达式两侧都调用对数，则得到：

$$\log f = \log c - s \log r$$

所以如果你以 $\log r$  为横轴给 $\log f$  绘图，则会得到斜率为 $-s$ ，截距为 $\log c$  的直线。

编写一个程序，从文件中读入文本，计算单词词频，并按照词频的降序，每一行打印出一个单词，以及 $\log f$  和 $\log r$ 。使用你喜欢的制图程序将结果以图表形式展现出来，并检查它是否为直线。你能估计 $s$  的值吗？

解答：<http://thinkpython.com/code/zipf.py>。要绘制图表，你可能需要安装matplotlib（参见 <http://matplotlib.sourceforge.net/>）。

# 第14章 文件

## 14.1 持久化

我们现在见过的程序都是瞬态的，因为它们会在短暂的时间里运行出一些输出，但当运行结束后，它们的数据会消失。如果再次运行程序，它会再次全新地开始。

也有些程序是持久化的：它们会运行很长一段时间（或者一直运行）；它们会至少存储一部分数据到永久存储（例如，硬盘）中；而且如果它们被关闭重启后，会接着从上次离开的状态继续。

持久化程序的例子包括操作系统，它几乎运行在任何一台开启的电脑中，以及web服务器，它们通常持续运行，等待网络上连入的请求。

读写文本文件是程序维护数据最简单的方法之一。我们已经见过读取文本文件的程序；在本章中将会见到往文件写入的程序。

另一种办法是将程序的状态保存到数据库中。本章中我们会介绍一个简单的数据库，以及一个模块，`pickle`，用来简化程序数据的存储。

## 14.2 读和写

文本文件是存储在诸如硬盘、闪存或光盘的永久媒介上的字符串序列。我们已经在9.1节中见过如何打开和读取一个文件。

要写入一个文件，你需要使用'w'模式作为第二个实参来打开它：

```
>>> fout = open('output.txt', 'w')
>>> print fout
<open file 'output.txt', mode 'w' at 0xb7eb2410>
```

如果文件已经存在，则使用写模式打开时会清除掉旧有数据并重新开始，所以请谨慎！ 如果文件不存在，则会新建一个。

`write`方法把数据写入到文件中。

```
>>> line1 = "This here's the wattle,\n"
>>> fout.write(line1)
```

同样地，文件对象会记录写到了哪里，所以如果你再次调用`write`，它会在结尾处添加新的数据。

```
>>> line2 = "the emblem of our land.\n"
>>> fout.write(line2)
```

当你写入完毕时，需要关闭文件。

```
>>> fout.close()
```

## 14.3 格式操作符

`write`的参数必须是字符串，所以若我们想要往文件中写入其他类型的值，必须将它们先转换为字符串。最容易的办法是使用`str`:

```
>>> x = 52  
>>> f.write(str(x))
```

另一个办法是使用格式操作符`%`。当用于整数时，`%`是求余操作符。但若第一个操作对象是字符串时，`%`则是格式操作符。

`%`的第一个操作对象是格式字符串，包括了一个或多个格式序列，由它们来指定第二个操作对象如何格式化。表达式的结果是一个字符串。

例如，格式序列`'%d'`意味着第二个操作数应该被格式化为整数（`d`表示“decimal”，即十进制）。

```
>>> camels = 42  
>>> '%d' % camels  
'42'
```

结果是字符串`'42'`，请不要将它和整数值42混淆。

格式序列可以出现在字符串的任意地方，所以你可以在一个句子中嵌入变量值：

```
>>> camels = 42  
>>> 'I have spotted %d camels.' % camels  
'I have spotted 42 camels.'
```

如果字符串中有多个格式序列，第二个操作对象就必须是元组。每个格式序列按顺序对应元组中的一个元素。

下面的例子使用`'%d'`格式化整数，`'%g'`格式化浮点数（别问为什么），以及`'%s'`格式化字符串：

```
>>> 'In %d years I have spotted %g %s.' % (3, 0.1, 'camels')
'In 3 years I have spotted 0.1 camels.'
```

元组中元素的个数必须和字符串中格式序列的个数一致。另外，元素的类型也要和格式序列一致：

```
>>> '%d %d %d' % (1, 2)
TypeError: not enough arguments for format string
>>> '%d' % 'dollars'
TypeError: illegal argument type for built-in operation
```

第一个例子中，元组中元素个数不够；第二个例子中，元素的类型不对。

格式操作符很强大，但可能难于使用。你可以在[docs.python.org\lib\typesseq-strings.html](https://docs.python.org/lib/typesseq-strings.html)阅读更多的相关信息。

## 14.4 文件名和路径

文件组织在目录（也称为文件夹）中。每个程序都有“当前目录”，它是大多数操作的默认目录。例如，当你打开一个文件用于读取时，Python默认在当前目录寻找它。

os模块提供了用于操作文件和目录的函数（os代表operating system，即操作系统）。os.getcwd返回当前目录的名称：

```
>>> import os  
>>> cwd = os.getcwd()  
>>> print cwd  
/home/dinsdale
```

cwd表示current working directory（即“当前工作目录”）。这个例子里的结果是/home/dinsdale，是名为dinsdale的用户的主目录。

类似于cwd这样用来定位文件的字符串被称为一个路径(path)。相对路径从当前目录开始；而绝对路径则从文件系统的顶层目录开始。

至今为止我们看到的路径都是简单的文件名，所以它们都是相对于当前目录的。你可以使用os.path.abspath来找寻文件的绝对路径：

```
>>> os.path.abspath('memo.txt')  
'/home/dinsdale/memo.txt'
```

os.path.exists检查一个文件或目录是否存在：

```
>>> os.path.exists('memo.txt')  
True
```

如果它存在，os.path.isdir检查它是否为目录：

```
>>> os.path.isdir('memo.txt')
False
>>> os.path.isdir('music')
True
```

类似地，`os.path.isfile`检查它是否为文件。

`os.listdir`返回指定目录中文件（以及其他目录）的列表：

```
>>> os.listdir(cwd)
['music', 'photos', 'memo.txt']
```

为了演示这些函数，下面的例子“走遍”一个目录，打印所有文件的名称，并对之中的子目录递归调用自己。

```
def walk(dirname):
    for name in os.listdir(dirname):
        path = os.path.join(dirname, name)

        if os.path.isfile(path):
            print path
        else:
            walk(path)
```

`os.path.join`接收一个目录和一个文件名称，并将它们拼接为一个完整的路径。

### 练习14-1

`os`模块提供了一个函数`walk`，和上面的例子作用类似，但功能更丰富。阅读文档，并使用它打印指定目录中文件的名称和它的子目录。

解答：<http://thinkpython.com/code/walk.py>。

## 14.5 捕获异常

当你尝试读取和写入文件时，很多东西都可能出错。如果你尝试打开一个不存在的文件，会得到一个`IOError`:

```
>>> fin = open('bad_file')
IOError: [Errno 2] No such file or directory: 'bad_file'
```

如果你没有权限访问一个文件:

```
>>> fout = open('/etc/passwd', 'w')
IOError: [Errno 13] Permission denied: '/etc/passwd'
```

而如果你尝试打开一个目录用于文件读取，会得到:

```
>>> fin = open('/home')
IOError: [Errno 21] Is a directory
```

要避免这些错误，可以使用类似`os.path.exists`和`os.path.isfile`的函数，但要检查所有的可能需要花费大量时间和代码（如果有“Errno 21”这样的指示，就说明至少有21种可能出错的地方）。

最好是直接去尝试——等发生问题时再去解决它们——这也正是`try`语句所做的事情。语法和`if`语句类似:

```
try:
    fin = open('bad_file')
    for line in fin:
        print line
    fin.close()
except:
    print 'Something went wrong.'
```

Python会先从`try`语句开始，如果一切顺利，则跳过`except`语句并继续执行。如果发生了异常，则跳出`try`语句，并执行`except`语句。

使用try语句处理异常的过程称为捕获一个异常。在这个例子  
里，except语句打印的错误信息并没有太多用处。总的来说，捕获  
异常给了你一个修补错误的机会，或者可以再次尝试，或者至少能够  
优雅地停止程序。

### 练习14-2

写一个函数sed，接收如下参数：一个模式字符串，一个替换用  
字符串，以及两个文件名；它应该读取第一个文件，并将内容写入到  
第二个文件（如果需要则新建它）。如果文件中任何地方出现了模式  
字符串，它应该被替换。

如果在打开、读取、写入或关闭文件的过程中遇到错误，你的程  
序应当能捕获异常，打印一个错误信息，并退出。解答：

<http://thinkpython.com/code/sed.py>。

## 14.6 数据库

数据库是一个有组织的用于存储数据的文件。大部分数据库都像字典一样组织数据，因为它们也将键映射到值上。最大的区别是数据库是保存在磁盘上（或者其他永久存储上）的，所以当程序结束时它也能持续存在。

模块anydbm提供了接口用于创建和更新数据库文件。作为示例，我将会创建一个数据库保存图片文件的标题。

打开一个数据库和打开其他类型的文件差不多：

```
>>> import anydbm  
>>> db = anydbm.open('captions.db', 'c')
```

模式'c'意味着数据库应当被创建，如果它不存在的话。调用的结果是一个数据库对象，（对大多数操作）可以当做字典来用。如果你创建一项，anydbm会更新数据库文件。

```
>>> db['cleese.png'] = 'Photo of John Cleese.'
```

当你访问数据库中的一项时，anydbm会读取文件：

```
>>> print db['cleese.png']  
Photo of John Cleese.
```

如果你对一个已经存在的键赋值，anydbm会替换旧值：

```
>>> db['cleese.png'] = 'Photo of John Cleese doing a silly walk.'  
>>> print db['cleese.png']  
Photo of John Cleese doing a silly walk.
```

很多字典方法，比如keys和items，对数据库对象也可以使用。这也包括使用for循环来迭代遍历。

```
for key in db:  
    print key
```

---

和其他文件一样，当操作结束时，你需要关闭数据库：

```
>>> db.close()
```

## 14.7 封存

anydbm的限制之一是键和值都必须是字符串。如果你尝试使用其他类型，则会出现错误。

pickle模块可以帮忙。它可以将几乎所有类型的对象转换为适合保存到数据库的字符串形式，并可以将字符串转换回来成为对象。

pickle.dumps接收一个对象作为参数，并返回它的字符串表达形式（dumps是“dump string”的简写，意即转储字符串）：

```
>>> import pickle  
>>> t = [1, 2, 3]  
>>> pickle.dumps(t)  
'(lp0\\nI1\\naI2\\naI3\\na.'
```

这个格式不适合人眼阅读；它是为了方便pickle模块的转换而设计的。pickle.loads（load string，即加载字符串）重新构造对象：

```
>>> t1 = [1, 2, 3]  
>>> s = pickle.dumps(t1)  
>>> t2 = pickle.loads(s)  
>>> print t2  
[1, 2, 3]
```

虽然新的对象和旧有对象的值相同，但（通常来说）它们不是同一个对象：

```
>>> t1 == t2  
True  
>>> t1 is t2  
False
```

也就是说，封存再解封，和复制对象效果相同。

你可以使用pickle向数据库存储非字符串的值。事实上，这个组合如此常用，以至于Python已经将它们封装起来成为一个模块，叫

做shelve。

### 练习14-3

如果你从[http://thinkpython.com/code/anagram\\_sets.py](http://thinkpython.com/code/anagram_sets.py)下载我对练习 12-4的解答，你会发现它创建一个字典，将一个排好序的字母串映射到可以由这些字母组成的单词的列表。例如，'opst'映射到列表

`['opts', 'post', 'pots', 'spot', 'stop', 'tops']`。

编写一个模块，导入anagram\_sets，并提供两个新函数：  
`store_anagrams`应当存储回文字典到一个“shelf”中；  
`read_anagrams`应当查询一个单词，并返回它的回文的列表。解答：[http://thinkpython.com/code/anagram\\_db.py](http://thinkpython.com/code/anagram_db.py)。

## 14.8 管道

大部分操作系统都提供了命令行接口，也称为字符界面（shell）。字符界面通常会提供命令来浏览文件系统和启动应用程序。例如，在Unix中，你可以使用`cd`来更换目录，使用`ls`来展示目录中的内容，以及打入`firefox`来启动浏览器。

任何在字符界面能启动的程序都可以在Python中通过一个管道（pipe）来启动。管道是代表一个正在运行的程序的对象。

例如，Unix命令`ls -l`（以长格式）展示当前目录的内容。你可以使用`os.popen`<sup>①</sup>来启动`ls`：

```
>>> cmd = 'ls -l'  
>>> fp = os.popen(cmd)
```

参数是一个字符串，它包含一个shell命令。返回值是一个和打开的文件差不多的对象。你可以使用`readline`来逐行读取`ls`进程的输出，或者使用`read`一次读取所有输出：

```
>>> res = fp.read()
```

当你完成时，可以像文件一样关闭这个管道：

```
>>> stat = fp.close()  
>>> print stat  
None
```

返回值是`ls`进程的最终状态；`None`代表它正常结束了（没有错误）。

例如，大部分Unix系统都提供了一个叫做`md5sum`的命令，它读取文件的内容并计算出一个“校验和”（checksum）。你可以在<http://en.wikipedia.org/wiki/Md5>阅读MD5的相关信息。这个命令提供了一个高效的方法，用来对比两个文件是否包含相同的内容。不同

的内容生成相同的校验和的几率极低（也就是，在宇宙崩溃之前不大可能发生）。

你可以在Python中使用管道来运行md5sum，并获得结果：

```
>>> filename = 'book.tex'
>>> cmd = 'md5sum ' + filename
>>> fp = os.popen(cmd)
>>> res = fp.read()
>>> stat = fp.close()
>>> print res
1e0033f0ed0656636de0d75144ba32e0  book.tex
>>> print stat
None
```

#### 练习14-4

在一个庞大的MP3集合中，有可能同一首歌有多个副本，保存在不同的目录中，或者文件名不同。这个练习的目的是搜索重复的歌。

1. 编写一个程序递归搜索目录及其所有的子目录，并返回所有指定后缀（如.mp3）的文件的完整路径的列表。提示：os.path提供了几个有用的方法来操纵文件和路径名称。
2. 要发现重复文件，你需要使用md5sum来计算每个文件的“校验和”。如果两个文件的校验和相同，它们很可能有相同的内容。
3. 你可以使用Unix命令diff来复审检验。

解答：[http://thinkpython.com/code/find\\_duplicates.py](http://thinkpython.com/code/find_duplicates.py)

## 14.9 编写模块

任何包含Python代码的文件都可以作为模块导入。例如，如果你有一个文件wc.py，其代码如下：

```
def linecount(filename):
    count = 0
    for line in open(filename):
        count += 1
    return count

print linecount('wc.py')
```

如果你运行这个程序，它会读取自身的内容，并打印出文件的行数，即7。你也可以像这样导入它：

```
>>> import wc
7
```

现在你有一个模块对象wc了：

```
>>> print wc
<module 'wc' from 'wc.py'>
```

它提供了一个函数linecount：

```
>>> wc.linecount('wc.py')
7
```

上述就是如何在Python中编写模块的方法。

这个例子唯一的问题是当你导入模块时，它会执行底部的测试代码。正常情况下，当你导入一个模块时，它会定义新的函数，但不会执行。

作为模块导入的程序，通常使用如下模式：

```
if __name__ == '__main__':
    print linecount('wc.py')
```

`__name__` 是一个内置变量，当程序启动时就会被设置。如果程序作为脚本执行，`__name__` 的值是 `__main__`；此时，测试代码会被执行。否则，如果程序作为模块被导入，则测试代码就被跳过了。

### 练习14-5

把这个例子输入到一个文件 `wc.py` 中，并将它作为一个脚本运行。然后运行 Python 解释器，并导入 `wc`。当模块被导入时，`__name__` 的值是什么？

#### 警告：

如果你导入一个已经被导入的模块，Python 什么都不做。它不会重新读取文件，即使文件已经修改。

如果你想要重载一个模块，可以使用内置函数 `reload`，但它也可能会有棘手的问题。所以最安全的办法是重启解释器，并再次导入模块。

## 14.10 调试

当你读取和写入文件时，可能会遇到和空白字符相关的问题。这些问题可能会很难调试，因为空格、制表符和换行符通常都是不可见的：

```
>>> s = '1 2\t 3\n 4'  
>>> print s  
1 2 3  
4
```

内置函数`repr`可以帮助。它接收任何对象作为参数，并返回对象的字符串表达形式。对于字符串来说，它使用反斜杠序列来展示空白字符：

```
>>> print repr(s)  
'1 2\t 3\n 4'
```

这样可以帮助调试。

另一个你可能遇到的问题是不同的系统使用不同的字符表示换行。有的系统使用一个换行符，即`\n`。另外的系统使用一个回车符，即`\r`。也有的系统两者都使用。如果你在不同的系统间移动文件，这些不一致之处可能会导致问题。

大多数系统都有程序可以将一种格式转换为另一种。你可以在<http://en.wikipedia.org/wiki/Newline>里找到它们（并阅读这个问题的更多信息）。或者，当然，你也可以自己写一个。

## 14.11 术语表

持久性 (persistent)：程序的一种属性，它会一直运行，并至少保存一部分数据在永久存储中。

格式操作符 (format operator)：一个操作符，即%，它接收一个格式字符串，以及一个元组，并生成字符串，其中包括了元组的各个依据格式字符串里指定的方式格式化的元素。

格式字符串 (format string)：一个字符串，被格式操作符所用，内部包含格式序列。

格式序列 (format sequence)：格式字符串中出现的字符序列，比如%d，它指定一个值如何格式化。

文本文件 (text file)：存储在类似硬盘这样的永久存储中的字符串序列。

目录 (directory)：有名称的文件集合。也称为文件夹。

路径 (path)：用来标定一个文件的字符串。

相对路径 (relative path)：从当前目录开始的路径。

绝对路径 (absolute path)：从文件系统的顶级目录开始的路径。

捕获 (catch)：使用try和except语句来阻止一个异常终止程序的行为。

数据库 (database)：一个文件，其内容组织类似于字典，将键映射到值。

## 14.12 练习

### 练习14-6

模块urllib提供了方法操纵URL，以及从网络上下载信息。下面的示例从thinkpython.com上下载并打印秘密信息：

```
import urllib

conn = urllib.urlopen('http://thinkpython.com/secret.html')
for line in conn:
    print line.strip()
```

运行这段代码，并听从你看到的指令。

解答：[http://thinkpython.com/code/zip\\_code.py](http://thinkpython.com/code/zip_code.py)。

---

① popen现在已经计划废止了，也就是说我们应当不再使用它，而是开始使用subprocess模块。但对于简单的情形，我发现subprocess过度复杂了。所以我仍然继续使用popen，直到它被完全废止。

# 第15章 类和对象

本章的代码示例可以从<http://thinkpython.com/code/Point1.py>下载；练习的解答在[http://thinkpython.com/code/Point1\\_soln.py](http://thinkpython.com/code/Point1_soln.py)。

# 15.1 用户定义类型

我们已经使用了很多Python的内置类型；现在我们要定义一个新类型。作为示例，我们将会新建一个类型Point，用来表示二维空间中的一个点。

在数学的表示法中，点通常使用括号中逗号分隔两个坐标表示。例如， $(0, 0)$ 表示原点，而 $(x, y)$ 表示一个在原点右侧 $x$ 单位，上方 $y$ 单位的点。

在Python中，有好几种方法可以表达点。

- 我们可以将两个坐标分别保存到变量x和y中。
- 我们可以将坐标作为列表或元组的元素存储。
- 我们可以新建一个类型用对象表达点。

新建一个类型比其他的方法（稍微）更复杂一些，但它的优点很快就会显现。

用户定义的类型也称为类（class）。类的定义如下所示：

```
class Point(object):
    """Represents a point in 2-D space."""
```

定义头表示新的类是一个Point，它是object的一种，而object是一个内置类型。

定义体是一个文档字符串，解释这个类的用途。你可以在类定义中定义变量和函数，我们会在后面回到这个话题。

定义一个叫做Point的类会创建一个类对象。

```
>>> print Point
<class '__main__.Point'>
```

因为Point是在程序顶层定义的，它的“全名”是`__main__.Point`。

类对象像一个创建对象的工厂。要新建一个Point对象，可以把Point当做函数来调用。

```
>>> blank = Point()
>>> print blank
<__main__.Point instance at 0xb7e9d3ac>
```

返回值是到一个Point对象的引用，我们将它赋值给变量blank。新建一个对象的过程称为**实例化**（instantiation），而对象是这个类的一个**实例**。

当你打印一个实例时，Python会告诉你它所属的类型，以及存储在内存中的位置（前缀0x表示后面的数字是十六进制的）。

## 15.2 属性

你可以使用句点表示法来给实例赋值：

```
>>> blank.x = 3.0  
>>> blank.y = 4.0
```

这个语法和从模块中选择变量的语法类似，例如`math.pi`或者`string.whitespace`。但在这个情况下，我们是将值赋给一个对象的有命名的元素。这些元素称为属性（attribute）。

作为名词时，“AT-trib-ute”发音的重音在第一个音节，这和作为动词的“a-TRIB-ute”不同。

下面的图表展示了这些赋值的结果。展示一个对象和其属性的状态图，我们称为对象图（object diagram）；参见图15-1。

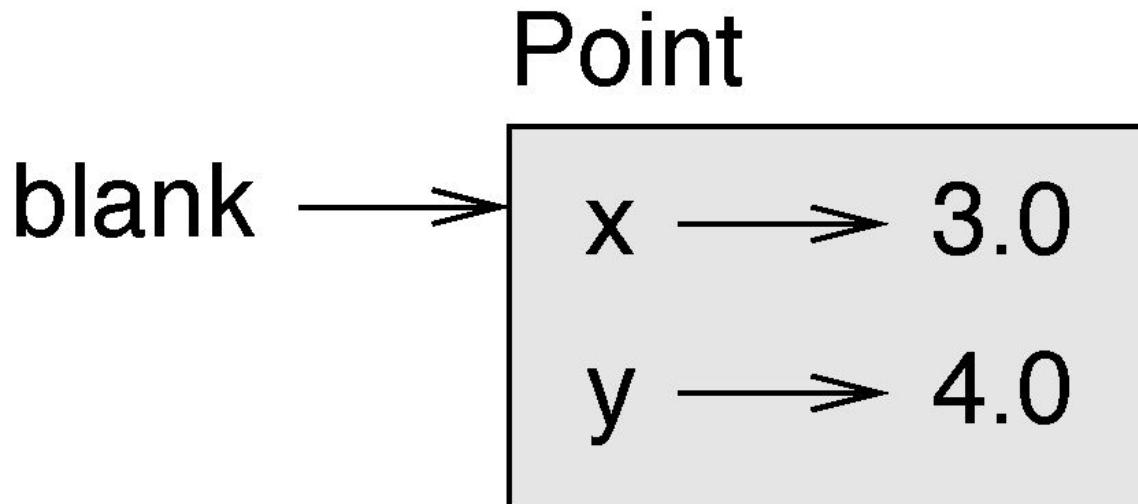


图15-1 对象图

变量`blank`引用向一个`Point`对象，它包含了两个属性。每个属性引用一个浮点数。

你可以使用相同的语法来读取一个属性的值：

```
>>> print blank.y  
4.0  
>>> x = blank.x
```

```
>>> print x  
3.0
```

表达式`blank.x`表示，“找到`blank`引用的对象，并取得它的`x`属性的值”。在这个例子中，我们将那个值赋值给一个变量`x`。变量`x`和属性`x`并不冲突。

你可以在任意表达式中使用句点表示法。例如：

```
>>> print '(%g, %g)' % (blank.x, blank.y)  
(3.0, 4.0)  
>>> distance = math.sqrt(blank.x**2 + blank.y**2)  
>>> print distance  
5.0
```

你可以将一个实例作为实参按通常的方式传递。例如：

```
def print_point(p):  
    print '(%g, %g)' % (p.x, p.y)
```

`print_point`接收一个点作为形参，并按照数学表达式展示它。你可以传入`blank`作为实参来调用它：

```
>>> print_point(blank)  
(3.0, 4.0)
```

在函数中，`p`是`blank`的一个别名，所以如果函数修改了`p`，则`blank`也会改变。

### 练习15-1

编写一个叫做`distance_between_points`的函数，接收两个`Point`对象作为形参，并返回它们之间的距离。

## 15.3 矩形

有时候对象应该有哪些属性非常明显，但也有时候需要你来做决定。例如，假设你在设计一个表达矩形的类。你会用什么属性来指定一个矩形的位置和尺寸呢？可以忽略角度；为了简单起见，假定矩形不是垂直的就是水平的。

最少有以下两种可能。

- 你可以指定一个矩形的一个角落（或者中心点）、宽度以及高度。
- 你可以指定两个相对的角落。

现在还很难说哪一种方案更好，所以作为示例，我们仅先实现第一个。

下面是这个类的定义：

```
class Rectangle(object):
    """Represents a rectangle.

    attributes: width, height, corner.
    """
```

文档字符串列出了属性：width和height是数字；corner是一个Point对象，用来指定左下角的顶点。

要表达一个矩形，你需要实例化一个Rectangle对象，并对其属性赋值：

```
box = Rectangle()
box.width = 100.0
box.height = 200.0
box.corner = Point()
box.corner.x = 0.0
box.corner.y = 0.0
```

表达式box.corner.x表示，“去往box引用的对象，并选择属性corner；接着去往那个对象，并选择属性x”。

图15-2展示了这个对象的状态。作为另一个对象的属性存在的对象是内嵌的。



图15-2 对象图

## 15.4 作为返回值的实例

函数可以返回实例。例如，`find_center`接收`Rectangle`对象作为参数，并返回一个`Point`对象，包含这个`Rectangle`的中心点的坐标：

```
def find_center(rect):
    p = Point()
    p.x = rect.corner.x + rect.width/2.0
    p.y = rect.corner.y + rect.height/2.0
    return p
```

下面是一个示例，传入`box`作为实参，并将结果的`Point`对象赋给变量`center`：

```
>>> center = find_center(box)
>>> print_point(center)
(50.0, 100.0)
```

## 15.5 对象是可变的

你可以通过给一个对象的某个属性赋值来修改它的状态。例如，要修改一个矩形的尺寸而保持它的位置不变，可以修改属性width和height的值：

```
box.width = box.width + 50  
box.height = box.width + 100
```

你也可以编写函数来修改对象。例如，grow\_rectangle接收一个Rectangle对象和两个数，dwidth和dheight，并把这些数加到矩形的宽度和高度上：

```
def grow_rectangle(rect, dwidth, dheight):  
    rect.width += dwidth  
    rect.height += dheight
```

下面是展示这个函数效果的示例：

```
>>> print box.width  
100.0  
>>> print box.height  
200.0  
>>> grow_rectangle(box, 50, 100)  
>>> print box.width  
150.0  
>>> print box.height  
300.0
```

在函数中，rect是box的别名，所以如果函数修改了rect，则box也改变。

### 练习15-2

编写一个叫做move\_rectangle的函数，接收一个Rectangle对象和两个数值，dx和dy。它应当通过添加dx到corner的x坐标和添加dy到corner的y坐标来改换矩形的位置。

## 15.6 复制

别名的使用有时候会让程序更难阅读，因为一个地方的修改可能会给其他地方带来意想不到的变化。要跟踪掌握所有引用到一个给定对象的变量非常困难。

使用别名的常用替代方案是复制对象。`copy`模块里有一个函数`copy`可以复制任何对象：

```
>>> p1 = Point()
>>> p1.x = 3.0
>>> p1.y = 4.0

>>> import copy
>>> p2 = copy.copy(p1)
```

`p1`和`p2`包含相同的数据，但是它们不是同一个`Point`对象。

```
>>> print_point(p1)
(3.0, 4.0)
>>> print_point(p2)
(3.0, 4.0)
>>> p1 is p2
False
>>> p1 == p2
False
```

和我们预料的相同，`is`操作符告诉我们`p1`和`p2`不是同一个对象。但你可能会预料`==`能得到`True`值，因为这两个点包含相同的数据。如果那样，你会失望地发现对于实例来说，`==`操作符的默认行为和`is`操作符相同；它会检查对象同一性，而不是对象相等性。这种行为可以改变——后面我们会看到如何去做。

如果你使用`copy.copy`复制一个`Rectangle`，你会发现它复制了`Rectangle`对象，但并不复制内嵌的`Point`对象。

```
>>> box2 = copy.copy(box)
>>> box2 is box
False
>>> box2.corner is box.corner
True
```

图15-3展示了这个操作的对象图。这个操作被称为浅复制，因为它复制对象和其包含的任何引用，但不复制内嵌对象。

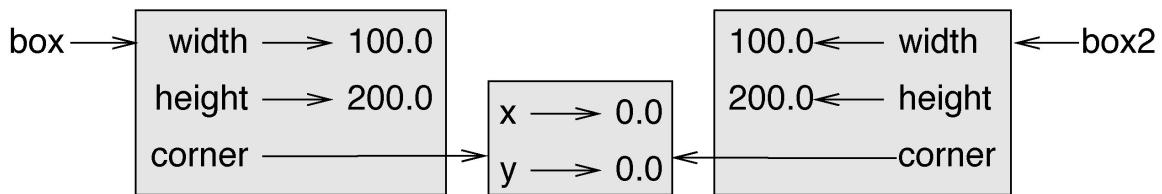


图15-3 对象图

对于大多数应用，这并不是你所想要的。在这个例子里，对一个 Rectangle对象调用`grow_rectangle`，并不会影响另外的对象，但对任何一个 Rectangle对象调用`move_rectangle`，都会影响到全部两个对象！这种行为既混乱不清，又容易导致错误。

幸运的是，`copy`模块还提供了一个方法`deepcopy`，它不但复制对象，还会复制对象中引用的对象，甚至它们引用的对象，依次类推。所以你并不会惊讶它为何被称为深复制（deep copy）。

```
>>> box3 = copy.deepcopy(box)
>>> box3 is box
False
>>> box3.corner is box.corner
False
```

box3和box是两个完全分开的对象。

### 练习15-3

编写`move_rectangle`的一个版本，它会新建并返回一个Rectangle对象，而不是直接修改旧有的对象。

## 15.7 调试

当你开始操作对象时，可能会遇到一些新的异常。

如果你试图访问一个并不存在的属性，会得到AttributeError：

```
>>> p = Point()  
>>> print p.z  
AttributeError: Point instance has no attribute 'z'
```

如果你不清楚一个对象是什么类型，可以问：

```
>>> type(p)  
<type '__main__.Point'>
```

如果你不确定一个对象是否拥有某个特定的属性，可以使用内置函数hasattr：

```
>>> hasattr(p, 'x')  
True  
>>> hasattr(p, 'z')  
False
```

第一个形参可以是任何对象；第二个形参是一个字符串，包含属性的名称。

## 15.8 术语表

类 (class)：一个用户定义的类型。类定义会新建一个类对象。

类对象 (class object)：一个包含用户定义类型的信息的对象。类对象可以用来创建该类型的实例。

实例 (instance)：属于某个类的一个对象。

属性 (attribute)：一个对象中关联的有命名的值。

内嵌对象 (embedded object)：作为一个对象的属性存储的对象。

浅复制 (shallow copy)：复制对象的内容，包括内嵌对象的引用；`copy`模块中的`copy`函数实现了这个功能。

深复制 (deep copy)：复制对象的内容，也包括内嵌对象，以及它们内嵌的对象，依此类推；`copy`模块中的`deepcopy`函数实现了这个功能。

对象图 (object diagram)：一个展示对象、它们的属性以及属性的值的图。

## 15.9 练习

### 练习15-4

Swampy（参见第4章）提供了一个叫做World的模块，其中定义了一个用户自定义类型，也叫做World。你可以这样导入它：

```
from swampy.World import World
```

下面的代码创建一个World对象并调用mainloop方法，它会等待用户的输入。

```
world = World()  
world.mainloop()
```

将会出现一个窗口，它有一个标题栏和一个空的方框。我们会使用这个窗口来绘制点（Points）、矩形（Rectangles）以及其他图形。在调用mainloop前添加下面的代码，并重新运行程序：

```
canvas = world.ca(width=500, height=500, background='white')  
bbox = [[-150, -100], [150, 100]]  
canvas.rectangle(bbox, outline='black', width=2, fill='green4')
```

你应该能见到一个绿色的矩形，边框是黑色的。第一行代码创建一个Canvas（画布）对象，它在窗口中展现为一个白色的方块。Canvas对象提供了像rectangle这样的方法用于绘制各类图形。

bbox是一个列表的列表，表示矩形的“边界框”。第一对坐标是矩形的左下角顶点；第二对坐标是矩形的右上角顶点。

你可以这样绘制一个圆形：

```
canvas.circle([-25, 0], 70, outline=None, fill='red')
```

第一个参数是一对坐标，表示圆心；第二个参数是其半径。

如果你将这行添加到程序中，结果会是一个类似孟加拉国国旗的图形（参见[http://en.wikipedia.org/wiki/Gallery\\_of\\_sovereign-state\\_flags](http://en.wikipedia.org/wiki/Gallery_of_sovereign-state_flags)）。

1. 编写一个函数`draw_rectangle`, 接收一个Canvas对象和一个Rectangle对象作为形参, 并在Canvas上绘制出这个Rectangle。
2. 在你的Rectangle对象上添加一个属性`color`, 并修改`draw_rectangle`函数, 让它能够根据`color`属性来给矩形填充颜色。
3. 编写一个函数`draw_point`, 接收一个Canvas对象和一个Point对象作为形参, 并在Canvas上绘制出这个Point。
4. 定义一个新的类Circle表示圆形, 给它定义合适的属性, 并实例化几个Circle对象。编写一个函数`draw_circle`在画布上绘制圆形。
5. 编写一个程序绘制捷克共和国的国旗。提示: 你可以这样绘制多边形:

```
points = [[-150,-100], [150, 100], [150, -100]]  
canvas.polygon(points, fill='blue')
```

我写了一个小程序, 能列出几种可用的颜色; 你可以从[http://thinkpython.com/code/color\\_list.py](http://thinkpython.com/code/color_list.py)下载它。

# 第16章 类和函数

本章的代码示例可以从<http://thinkpython.com/code/Time1.py>下载。

## 16.1 时间

作为用户定义类型的另一个例子，我们定义一个叫做Time的类，用于记录一天里的时间。类定义如下：

```
class Time(object):
    """Represents the time of day.

    attributes: hour, minute, second
    """
```

我们可以创建一个时间对象并给其属性小时数、分钟数和秒钟数赋值：

```
time = Time()
time.hour = 11
time.minute = 59
time.second = 30
```

时间对象的状态图参见图16-1。

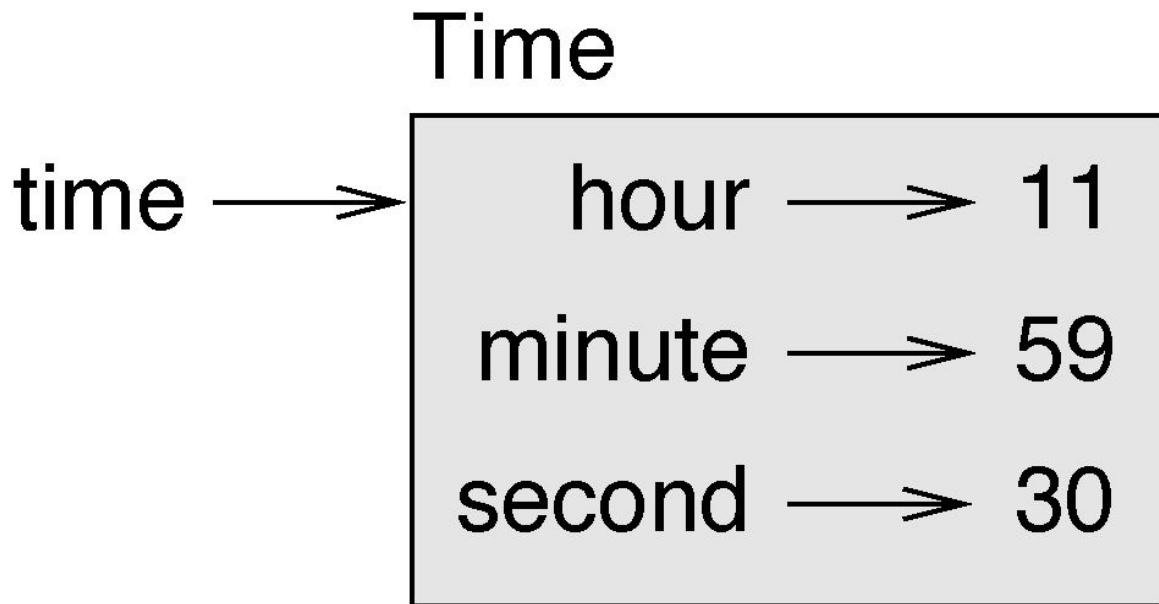


图16-1 对象图

练习16-1

编写一个叫做print\_time的函数，接收一个时间对象作为形参并以“时:分:秒”的格式打印它。提示：格式序列'%.2d'可以以最少两个字符打印一个整数，如果需要，它会在前面添加前缀0。

### 练习16-2

编写一个布尔函数is\_after，接收两个时间对象，t1和t2，并若t1在t2时间之后则返回True，否则返回False。挑战：不许使用if表达式。

## 16.2 纯函数

在下面几节中，我们会编写两个用来增加时间值的函数。它们展示了两种不同类型的函数：纯函数和修改器。它们也展示了我会称为原型和补丁（prototype and patch）的开发计划。这是一种对应复杂问题的方法，从一个简单的原型开始，并逐渐解决更多的复杂情况。

下面是add\_time的一个简单原型：

```
def add_time(t1, t2):
    sum = Time()
    sum.hour = t1.hour + t2.hour
    sum.minute = t1.minute + t2.minute
    sum.second = t1.second + t2.second
    return sum
```

这个函数创建一个新的时间对象，初始化它的属性，并返回这个新对象的一个引用。这被称为一个纯函数，因为它除了返回一个值之外，并不修改作为实参传入的任何对象，也没有任何如显示值或获得用户输入之类的副作用。

为了测试这个函数，我将创建两个时间对象：start存放一个电影（如*Monty Python and the Holy Grail*）的开始时间，以及duration，存放电影的播放时间，在这里是1小时35分钟。

add\_time计算出电影何时结束。

```
>>> start = Time()
>>> start.hour = 9
>>> start.minute = 45
>>> start.second = 0

>>> duration = Time()
>>> duration.hour = 1
>>> duration.minute = 35
>>> duration.second = 0

>>> done = add_time(start, duration)
>>> print_time(done)
10:80:00
```

结果，10:80:00，可能并不是你所期望的。问题在于这个函数并没有处理好秒数或者分钟数超过60的情况。当此发生时，我们需要将多余的秒

数“进位”到分钟数，将多余的分钟数“进位”到小时数。

下面是一个改进的版本：

```
def add_time(t1, t2):
    sum = Time()
    sum.hour = t1.hour + t2.hour
    sum.minute = t1.minute + t2.minute
    sum.second = t1.second + t2.second

    if sum.second >= 60:
        sum.second -= 60
        sum.minute += 1

    if sum.minute >= 60:
        sum.minute -= 60
        sum.hour += 1

    return sum
```

虽然这个函数是正确的，它已经开始变大了。我们会在后面看到一个更短的版本。

## 16.3 修改器

有时候用函数修改传入的参数对象是很有用的。在这种情况下，修改对调用者是可见的。这样工作的函数称为修改器（modifier）。

函数increment给一个时间对象增加指定的秒数，可以自然地写为一个修改器。下面是一个初稿：

```
def increment(time, seconds):
    time.second += seconds

    if time.second >= 60:
        time.second -= 60
        time.minute += 1

    if time.minute >= 60:
        time.minute -= 60
        time.hour += 1
```

第一行进行基础操作；后面的代码处理我们前面看到的特殊情况。

这个函数正确吗？如果形参seconds比60大很多，会发生什么？在那种情况下，只进位一次是不够的；我们需要重复进位，直到time.second比60小。一个办法是使用while语句替代if语句。那样会让函数变正确，但并不很高效。

### 练习16-3

编写正确的increment版本，并不包含任何循环。

任何可以使用修改器做到的功能都可以使用纯函数实现。事实上，有的编程语言只允许使用纯函数。有证据表明使用纯函数的程序比使用修改器的程序开发更快，错误更少。但有时候修改器还是很方便的，并且函数式程序的运行效率不那么高。

总的来说，我推荐你只要合理的时候，都尽量编写纯函数，而只在有绝对说服力的原因时才使用修改器。这种方法可以称作函数式编程风格。

### 练习16-4

编写一个increment的纯函数版本，创建并返回一个新的时间对象，而不是修改参数。

## 16.4 原型和计划

刚才我展示的开发计划称为“原型和补丁”。对每个函数，我编写一个可以进行基本计算的原型，再测试它，从中发现错误并打补丁。

这个方法在你对问题的理解并不深入时尤其有效。但增量地修正可能会导致代码过度复杂，因为它们需要处理很多特殊情况，并且也不够可靠——因为很难知道你是否已经找到了所有错误。

另一种方法是有规划开发。对问题有更高阶的理解能够让编程简单得多。在这里，更深入的理解是，时间对象实际上是六十进制数里的3位数（参见<http://en.wikipedia.org/wiki/Sexagesimal>）！second属性是“个位数”，minute属性是“60位数”，而hour属性是“360位数”。

当我们编写add\_time和increment时，我们实际上是在六十进制上进行加减，因此才需要从一位进位到另一位。

这个观察让我们可以考虑整个问题的另一种解决方法——我们可以将时间对象转换为整数，并利用计算机知道如何做整数运算的事实。

下面是一个将时间对象转换为整数的函数：

```
def time_to_int(time):
    minutes = time.hour * 60 + time.minute
    seconds = minutes * 60 + time.second
    return seconds
```

而下面是一个将整数转换回时间对象的函数（记着divmod函数将第一个参数除以第二个参数，并以元组的形式返回商和余数）。

```
def int_to_time(seconds):
    time = Time()
    minutes, time.second = divmod(seconds, 60)
    time.hour, time.minute = divmod(minutes, 60)
    return time
```

你可能需要思考一下，并运行一些测试，来说服自己这些函数是正确的。一种测试它们的方法是对很多x值检查  
time\_to\_int(int\_to\_time(x)) == x。这是一致性检验的一个例子。

一旦你确认它们是正确的，就可以使用它们重写add\_time：

```
def add_time(t1, t2):
    seconds = time_to_int(t1) + time_to_int(t2)
    return int_to_time(seconds)
```

这个版本比最初版本短得多，并且也很容易检验。

### 练习16-5

使用time\_to\_int和int\_to\_time重写increment函数。

从某个角度看，在六十进制和十进制之间来回转换比只处理时间更难。进制转换更加抽象；我们对时间值的直觉更好。

但如果我们将时间看做六十进制数，并做好了编写转换函数（time\_to\_int和int\_to\_time）的前期投入，就能得到一个更短，更可读，也更可靠的函数。

它也让我们今后更容易添加功能。例如，假设将两个时间对象相减来获得它们之间的时间间隔。简单的做法是使用借位实现减法。而使用转换函数则更简单，且更容易正确。

讽刺的是，有时候把一个问题弄得更难（或者更通用）反而会让它更简单（因为会有更少的特殊情况以及更少的出错机会）。

## 16.5 调试

一个时间对象当minute和second的值在0到60之间（包含0但不包含60）以及hour是正值时，是合法的。hour和minute应当是整数值，但我们也许需要允许second拥有小数值。

这些需求称为不变式，因为它们应当总是为真。换句话说，如果它们不为真，则一定有什么地方出错了。

编写代码来检查不变式可以帮助你探测错误并找寻它们的根源。例如，你可以写一个像valid\_time这样的函数，接收时间对象，并当它违反了一个不变式时，返回False：

```
def valid_time(time):
    if time.hour < 0 or time.minute < 0 or time.second < 0:
        return False
    if time.minute >= 60 or time.second >= 60:
        return False
    return True
```

接着在每个函数的开头，你可以检查参数，确保它们是有效的：

```
def add_time(t1, t2):
    if not valid_time(t1) or not valid_time(t2):
        raise ValueError, 'invalid Time object in add_time'
    seconds = time_to_int(t1) + time_to_int(t2)
    return int_to_time(seconds)
```

或者你可以使用一个assert语句。它会检查一个给定的不变式，并当检查失败时抛出异常：

```
def add_time(t1, t2):
    assert valid_time(t1) and valid_time(t2)
    seconds = time_to_int(t1) + time_to_int(t2)
    return int_to_time(seconds)
```

assert语句很有用，因为它们区分了处理普通条件的代码和检查错误的代码。

## 16.6 术语表

原型和补丁（prototype and patch）：一种开发计划模式，先编写程序的粗略原型，并测试，在找到错误时更正。

有规划开发（planned development）：一种开发计划模式，先对问题有了高阶的深入理解，并且比增量开发或者原型开发有更多的规划。

纯函数（pure function）：不修改任何形参对象的函数。大部分纯函数都有返回值。

修改器（modifier）：修改一个或多个形参对象的函数。大部分修改器都不返回值。

函数式编程风格（functional programming style）：一种编程设计风格，其中大部分函数都是纯函数。

不变式（invariant）：在程序的执行过程中应当总是为真的条件。

## 16.7 练习

本章中的代码示例可以从<http://thinkpython.com/code/Time1.py>下载；这些练习的解答可以从[http://thinkpython.com/code/Time1\\_soln.py](http://thinkpython.com/code/Time1_soln.py)下载。

### 练习16-6

编写一个函数`mul_time`接收一个时间对象以及一个整数，返回一个新的时间对象，包含原始时间和整数的乘积。

然后使用`mul_time`来编写一个函数，接收一个时间对象表示一场赛车的结束时间，以及一个表示距离的数字，并返回一个时间对象表达平均节奏（每英里花费的时间）。

### 练习16-7

`datetime`模块提供了`date`和`time`对象，和本章中的日期与时间对象类似，但它们提供了更丰富的方法和操作符。在<http://docs.python.org/lib/datetime-date.html>阅读相关文档。

1. 使用`datetime`模块来编写一个程序获取当前日期并打印出今天是周几。
2. 编写一个程序接收生日作为输入，并打印出用户的年龄，以及到他们下一次生日还需要的天数、小时数、分钟数和秒数。
3. 对于生于不同天的两个人，总有一天，一个人的年龄是另一个人的两倍。我们称这是他们的“双倍日”。编写一个程序接收两个生日，并计算出它们的“双倍日”。
4. 再增加一点挑战，编写一个更通用的版本，计算一个人比另一个人大 $n$ 倍的日子。

# 第17章 类和方法

本章的代码示例可以从<http://thinkpython.com/code/Time2.py>下载。

## 17.1 面向对象特性

Python是一门面向对象编程语言，它提供了一些语言特性来支持面向对象编程。

定义面向对象编程并不是件容易事，但我们已经见过它的一些特征了。

- 程序由对象定义和函数定义组成，并且大部分的计算都是用对象的操作来表达的。
- 每个对象定义对应真实世界的某些对象或概念，并且操作那个对象的函数对应了真实世界中对象之间交互的方式。

例如，第16章中定义的Time类对应于人们记录一天中的时间的方式，而其中我们定义的函数对应于人们平时处理时间所做的事情。类似地，Point和Rectangle类对应于数学中点和矩形的概念。

至今为止，我们还没有利用上Python所提供的面向对象编程特性。严格地说，这些特性并不是必需的；它们中大部分都是我们已经做过的事情的另一种选择方案。但在很多情况下，这种方案更简洁，更能准确地表达程序的结构。

例如，在Time程序中，类定义和接着的函数定义并没有明显的关联。稍加观察，很明显每个函数都至少接收一个时间对象作为参数。这种现象是方法的由来；一个方法，即是和某个特定类相关联的函数。我们已经见过字符串、列表、字典和元组的方法。本章中，我们会为用户定义类型定义方法。

方法和函数在语义上是一样的，但在语法上有两个区别。

- 方法定义写在类定义之中，更明确的表示类和方法的关联。
- 调用方法和调用函数的语法形式不同。

在接下来几节中，我们会将前两章中定义的函数转换为方法。这种转换是纯机械式的；你可以简单地依照一系列步骤完成它。如果你能够轻松地在方法和函数之间转换，也就能够在任何情况下选择最适合的形式了。

## 17.2 打印对象

在第16章中，我们在练习16-1定义了一个叫做Time的类，你写过一个print\_time函数：

```
class Time(object):
    """Represents the time of day."""

def print_time(time):
    print '%.2d:%.2d:%.2d' % (time.hour, time.minute, time.second)
```

要调用这个函数，需要传入一个时间对象作为实参：

```
>>> start = Time()
>>> start.hour = 9
>>> start.minute = 45
>>> start.second = 00
>>> print_time(start)
09:45:00
```

要把print\_time转换为方法，我们只需要将函数定义移动到类定义之内即可。注意缩进的改变。

```
class Time(object):
    def print_time(time):
        print '%.2d:%.2d:%.2d' % (time.hour, time.minute, time.second)
```

现在有两种方式可以调用print\_time。第一种（更少见的）方式使用函数调用语法：

```
>>> Time.print_time(start)
09:45:00
```

在这里的句点表示法中，Time是类的名称，而print\_time是方法的名称。start是作为参数传入的。

另一种（更简洁的）方式是使用方法调用语法：

```
>>> start.print_time()
09:45:00
```

在这里的句点表示法中，`print_time`（又一次）是方法的名称，而`start`是调用这个方法的对象，也称为主体（subject）。和一句话中主语用来表示这句话是关于什么东西一样，方法调用的主体表示这个方法是关于哪个对象的。

在方法之中，主体会被赋值到第一个形参上，所以本例中`start`被赋值到`time`上。

依惯例来，方法的第一个形参通常叫做`self`，所以`print_time`通常写成这样的形式：

```
class Time(object):
    def print_time(self):
        print '%.2d:%.2d:%.2d' % (self.hour, self.minute, self.second)
```

这种惯例的原因是一个隐喻。

- 函数调用的语法，`print_time(start)`，暗示函数是活动主体。它仿佛在说：“喂，`print_time`！这里是一个让你打印的对象。”
- 在面向对象编程中，对象是活动主体。类似`start.print_time()`的方法调用相当于说：“喂，`start`！请打印你自己。”

这种视角的改变可能变得更礼貌，但它同时也更有用。这一点却不容易看出来。在我们已经见过的例子中，它也许并不更有用。但有时候将函数的责任转到对象上，使得我们能够编写功能更丰富的函数，也使得代码的维护和复用更容易。

### 练习17-1

将16.4节中的函数`time_to_int`转换为方法。而将`int_to_time`转换为方法也许并不合适；你会在什么对象上调用它呢？

## 17.3 另一个示例

下面是函数increment（见16.3节）的另一个版本，重写成了方法：

```
# inside class Time:  
  
    def increment(self, seconds):  
        seconds += self.time_to_int()  
        return int_to_time(seconds)
```

这个版本假设time\_to\_int已经写成了方法，如练习17-1所示。另外，注意它是一个纯函数，而不是一个修改器。

下面展示如何调用increment：

```
>>> start.print_time()  
09:45:00  
>>> end = start.increment(1337)  
>>> end.print_time()  
10:07:17
```

主体start赋值给第一个形参self，实参1337，赋值给第二个形参seconds。

这种机制有时也会带来困惑，尤其在当程序出错的时候。例如，如果使用两个实参数调用increment，则会得到：

```
>>> end = start.increment(1337, 460)  
TypeError: increment() takes exactly 2 arguments (3 given)
```

错误信息初看起来似乎很令人困惑，因为括号里只有两个实参。但调用的主体也被看作一个实参，所以其实总共有3个。

## 17.4 一个更复杂的示例

函数`is_after`（见练习16-2）稍微更复杂一些，因为它接收两个时间对象作为形参。这种情形下，依惯例，第一个形参命名为`self`，而第二个形参命名为`other`：

```
# inside class Time:  
  
    def is_after(self, other):  
        return self.time_to_int() > other.time_to_int()
```

要使用这个方法，你需要在一个对象上调用它，并传入另一个对象作为实参：

```
>>> end.is_after(start)  
True
```

这种语法的一个好处是，阅读起来几乎和英语一样：“`end is after start?`”。

## 17.5 `init`方法

`init`方法（即“`initialization`”的简写）是一个特殊方法，当对象初始化时会被调用。它的全名是`__init__`（两个下划线，接着是`init`，再接着两个下划线）。`Time`类的`init`方法可能如下所示：

```
# inside class Time:  
  
    def __init__(self, hour=0, minute=0, second=0):  
        self.hour = hour  
        self.minute = minute  
        self.second = second
```

`__init__`的形参和类的属性名称常常是相同的。

语句`self.hour = hour`将形参`hour`的值存储为`self`的一个属性。

形参是可选的，所以当你不使用任何实参调用`Time`时，会得到默认值：

```
>>> time = Time()  
>>> time.print_time()  
00:00:00
```

如果你提供一个实参，它会覆盖`hour`：

```
>>> time = Time(9)  
>>> time.print_time()  
09:00:00
```

如果你提供两个实参，它会覆盖`hour`和`minute`：

```
>>> time = Time(9, 45)  
>>> time.print_time()  
09:45:00
```

如果你提供3个实参，它们会覆盖全部3个默认值。

练习17-2

为Point类编写一个init方法，接收x和y作为可选形参，并将它们的值赋到对应的属性上。

## 17.6 `__str__`方法

`str` 和 `__init__` 类似，是一个特殊方法，它用来返回对象的字符串表达形式。

例如，下面是一个时间对象的`str`方法：

```
# inside class Time:  
  
    def __str__(self):  
        return '%.2d:%.2d:%.2d' % (self.hour, self.minute, self.second)
```

当你打印对象时，Python会调用`str`方法。

```
>>> time = Time(9, 45)  
>>> print time  
09:45:00
```

当我编写一个新类时，我总是开始先写`__init__`，以便初始化对象，然后会写`__str__`，以便调试。

### 练习17-3

为`Point`类编写一个`str`方法。创建一个`Point`对象并打印它。

## 17.7 操作符重载

通过定义其他的特殊方法，你可以为用户定义类型的各种操作符指定行为。例如，如果你为Time类定义一个`__add__`方法，则可以在时间对象上使用+操作符。

下面是这个方法的定义：

```
# inside class Time:  
  
def __add__(self, other):  
    seconds = self.time_to_int() + other.time_to_int()  
    return int_to_time(seconds)
```

而下面是如何使用它：

```
>>> start = Time(9, 45)  
>>> duration = Time(1, 35)  
>>> print start + duration  
11:20:00
```

当你对时间对象应用+操作符时，Python会调用`__add__`。当你打印结果时，Python会调用`__str__`。所以幕后其实发生了很多事情！

修改操作符的行为以便它能够作用于用户定义类型，这个过程称为操作符重载。对每一个操作符，Python都提供了一个对应的特殊方法，比如`__add__`。更多细节，可以参见<http://docs.python.org/ref/specialnames.html>。

练习17-4

为Point类编写一个add函数。

## 17.8 基于类型的分发

在前面一节中我们将两个时间对象相加，但你也可能会想要将一个时间对象加上一个整数。接下来是`__add__`的一个版本，检查`other`的类型，并调用`add_time`或`increment`:

```
# inside class Time:

    def __add__(self, other):
        if isinstance(other, Time):
            return self.add_time(other)
        else:
            return self.increment(other)

    def add_time(self, other):
        seconds = self.time_to_int() + other.time_to_int()
        return int_to_time(seconds)

    def increment(self, seconds):
        seconds += self.time_to_int()
        return int_to_time(seconds)
```

内置函数`isinstance`接收一个值与一个类对象，并当此值是此类的一个实例时返回`True`。

如果`other`是一个时间对象，`__add__`会调用`add_time`。否则它认为实参是整数，并调用`increment`。这个操作称为**基于类型的分发**(type-based dispatch)，因为它根据形参的类型，将计算分发到不同的方法上。

下面的示例，使用不同类型的实参调用`+`操作符：

```
>>> start = Time(9, 45)
>>> duration = Time(1, 35)
>>> print start + duration
11:20:00
>>> print start + 1337
10:07:17
```

不幸的是，这个加法的实现并不满足交换律。如果整数是第一个操作数，则会得到：

```
>>> print 1337 + start
TypeError: unsupported operand type(s) for +: 'int' and 'instance'
```

问题在于，这里和之前询问一个时间对象加上一个整数不同，Python在询问一个整数去加上一个时间对象，而它并不知道如何去做。但这个问题也有一个聪明的解决方案：特别方法`__radd__`，意即“右加法”（right-side add）。当时间对象出现在+号的右侧时，会调用这个方法。下面是它的定义：

```
# inside class Time:

    def __radd__(self, other):
        return self.__add__(other)
```

而下面是如何使用：

```
>>> print 1337 + start
10:07:17
```

### 练习17-5

为Point类编写一个add方法，可以接收一个Point对象或者一个元组。

- 如果第二个操作对象是一个Point对象，则方法应该返回一个新的Point对象，其x坐标是两个操作对象的x坐标的和，y坐标也是类似。
- 如果第二个操作对象是一个元组，方法则将第一个元素和x坐标相加，将第二个元素和y坐标相加，并返回一个包含相加结果的新Point对象。

## 17.9 多态

当需要时，基于类型的分发很有用，但（幸运的是）我们并不总是需要它。通常你可以编写函数处理不同类型的参数来避免它。

我们编写的很多处理字符串的函数，实际上对任何序列都可以用。例如，在11.1节中，我们使用`histogram`来记录单词中每个字母出现的次数。

```
def histogram(s):
    d = dict()
    for c in s:
        if c not in d:
            d[c] = 1
        else:
            d[c] = d[c]+1
    return d
```

这个函数对列表、元组甚至是字典都可用，只要s的元素是可散列的、因而可以用做d的键即可。

```
>>> t = ['spam', 'egg', 'spam', 'spam', 'bacon', 'spam']
>>> histogram(t)
{'bacon': 1, 'egg': 1, 'spam': 4}
```

可以处理多个类型的函数称为**多态**（polymorphic）。多态可以促进代码复用。例如，用来计算一个序列所有元素的和的内置函数`sum`，对所有其元素支持加法的序列都可用。

由于时间对象提供了`add`方法，它们也可以使用`sum`:

```
>>> t1 = Time(7, 43)
>>> t2 = Time(7, 41)
>>> t3 = Time(7, 37)
>>> total = sum([t1, t2, t3])
>>> print total
23:01:00
```

总的来说，如果函数内部所有的操作都支持某种类型，那么这个函数就可以用于那种类型。

最好的多态是，当你发现一个写好的函数竟可以用于从未预料过的类型。

## 17.10 调试

在程序运行的任何时刻，往对象上添加属性都是合法的，但如果你遵守更严格的类型理论，让对象拥有相同的类型却有不同的属性组，是非常奇怪的事情。通常来说，在`init`方法中初始化对象的全部属性是个好习惯。

如果你并不清楚一个对象是否拥有某个属性，可以使用内置函数`hasattr`（参见15.7节）。

另一种访问一个对象的属性的方法是通过特别属性`__dict__`，它是一个映射，将属性名称（字符串形式）映射到属性值：

```
>>> p = Point(3, 4)
>>> print p.__dict__
{'y': 4, 'x': 3}
```

为了调试方便，你可能会发现将这个函数放在手边是很有用的：

```
def print_attributes(obj):
    for attr in obj.__dict__:
        print attr, getattr(obj, attr)
```

`print_attributes`遍历对象的属性字典，并打印出每个属性的名称和相应的值。

内置函数`getattr`接收一个对象以及一个属性名称（字符串形式）并返回属性的值。

## 17.11 接口和实现

面向对象设计的目标之一是提高软件的可维护性，也就是说，当系统的其他部分改变时，程序还能够保持正确运行，并且能够修改程序来适应新的需求。

将接口和实现分离的设计理念，可以帮助我们更容易达到这个目标。对于对象来说，那意味着类所提供的方法应该不依赖于其属性的表达形式。

例如，在本章中我们开发了一个类来表示一天中的时间。这个类提供的方法包括`time_to_int`、`is_after`以及`add_time`。

我们可以使用几种不同的方式来实现这些方法。实现的细节依赖于我们如何表达时间概念。在本章中，时间对象的属性是`hour`、`minute`及`second`。

用另一种方案，我们可以将这些属性替换成单一的一个整数，表示从凌晨开始到现在的秒数。这种实现可能会让一些方法，比如`is_after`，更容易实现，但也会让另一些方法更难实现。

当你部署一个新类时，可能会发现更好的实现。如果程序中其他部分使用到你的类，则修改接口会非常消耗时间，并且容易产生错误。

但如果你很谨慎小心地设计接口，则可以在不修改接口的情况下修改实现，这样程序的其他部分则不需要跟着修改。

将接口和实现分开，意味着你需要将属性隐藏起来。程序其他部分的代码（类定义之外的），应当使用方法来读写对象的状态。它们不应当直接访问属性。这个原则称为信息隐藏（information hiding）；参见[http://en.wikipedia.org/wiki/Information\\_hiding](http://en.wikipedia.org/wiki/Information_hiding)。

### 练习17-6

从<http://thinkpython.com/code/Time2.py>下载本章的代码。修改`Time`的属性，用单一的整数表示凌晨到现在的秒数。然后修改方法（以及函数`int_to_time`）来适应新的实现。你应当不需要修改`main`函数

中的测试代码。当你完成时，输出应当和之前一样。解答：  
[http://thinkpython.com/code/Time2\\_soln.py](http://thinkpython.com/code/Time2_soln.py)。

## 17.12 术语表

面向对象语言 (object-oriented language)：一种提供诸如用户定义类型、方法语法之类的语言特性，以方便面向对象编程的语言。

面向对象编程 (object-oriented programming)：一种编程风格，数据和修改数据的操作组织成类和方法的形式。

方法 (method)：在类定义之内定义的函数，在类的实例上调用。

主体 (subject)：调用方法所在的对象。

操作符重载 (operator overloading)：修改一个类似+号这样的操作符的行为，使之可以用于用户定义类型。

基于类型的分发 (type-based dispatch)：一种编程模式，检查操作对象的类型，并对不同类型调用不同的函数。

多态 (polymorphic)：函数的一种属性，可以处理多种类型的参数。

信息隐藏 (information hiding)：对象提供的接口不应当依赖于其实现，特别是其属性的表达形式的原则。

## 17.13 练习

### 练习17-7

这个练习提醒你关于Python的一种最常见且最难查找的错误的故事。

编写一个叫做Kangaroo（袋鼠）的类，有如下方法。

1. 一个`__init__`方法，将属性`pouch_contents`（口袋中的东西）初始化为一个空列表。
2. 一个`put_in_pouch`方法，接收任何类型的对象，并将它添加到`pouch_contents`中。
3. 一个`__str__`方法，返回Kangaroo对象以及口袋中的内容的字符串表达形式。

创建两个Kangaroo对象，将它们赋值到变量`kanga`和`roo`，并将`roo`添加到`kanga`的口袋中。

下载<http://thinkpython.com/code/BadKangaroo.py>，它包含了前面问题的解答，但里面有一个很大很丑陋的bug。找出并修复这个bug。

如果你遇到阻碍，可以下载  
<http://thinkpython.com/code/GoodKangaroo.py>，它解释了问题的原因，并提供了一个解决方案。

### 练习17-8

Visual是一个Python模块，提供3-D图形操作的功能。它并不总是包含在Python安装包中，所以你可能需要从软件源中安装它，如果没有的话，可以从<http://vpython.org>下载。

下面的示例创建一个3D空间，长、宽、高都是256单位，而“中心”点设置为点(128, 128, 128)。然后它绘制一个蓝色的球体。

```
from visual import *
```

```
scene.range = (256, 256, 256)
scene.center = (128, 128, 128)

color = (0.1, 0.1, 0.9)      # 几乎完全是蓝色
sphere(pos=scene.center, radius=128, color=color)
```

color是一个RGB元组，每个元素是红-绿-蓝（Red-Green-Blue）颜色的级别，从0.0到1.0之间。参见[http://en.wikipedia.org/wiki/RGB\\_color\\_model](http://en.wikipedia.org/wiki/RGB_color_model)。

如果你运行这段代码，会看到黑色背景和一个蓝色球体。如果你上下拖拽底部的按钮，可以放大或缩小。你也可以拖拽右侧的按钮来旋转整个场景，但因为这个世界中只有一个球体，所以基本看不出区别。

下面的循环创建一个由球体组成的立方体。

```
t = range(0, 256, 51)
for x in t:
    for y in t:
        for z in t:
            pos = x,y,z
            sphere(pos=pos, radius=10, color=color)
```

1. 将这段代码放到一个脚本中，并保证它能够正确运行。
2. 修改程序，让立方体中的每个球体都有对应于其位置的RGB值。注意坐标的范围是0-255，但是RGB元组的范围是0.0-1.0。
3. 下载[http://thinkpython.com/code/color\\_list.py](http://thinkpython.com/code/color_list.py)并使用函数read\_colors来生成一系列你的系统可用的颜色，它们的名字，以及它们的RGB值。对每个命名的颜色，在对应其RGB值的位置画一个球体。

你可以在[http://thinkpython.com/code/color\\_space.py](http://thinkpython.com/code/color_space.py)看到我的解答。

# 第18章 继承

本章中我会使用几个类来表达扑克牌、牌组以及扑克牌型。如果你不玩扑克，可以在<http://en.wikipedia.org/wiki/Poker>里阅读相关介绍，但其实并不必要；我会告诉你练习中所需知道的东西。本章的代码示例可以从<http://thinkpython.com/code/Card.py>下载。

如果你不熟悉英美式牌类游戏，可以在[http://en.wikipedia.org/wiki/Playing\\_cards](http://en.wikipedia.org/wiki/Playing_cards)阅读相关信息。

## 18.1 卡片对象

一副牌里有52张牌，共有4个花色，每种花色13张，大小各不相同。花色有黑桃（Spade）、红桃（Heart）、方片（Diamond）和草花（Club）（在桥牌中，这几个花色是降序排列的）。每种花色的13张牌分别为：Ace、2、3、4、5、6、7、8、9、10、Jack、Queen和King。根据你玩的不同游戏，Ace可能比King大，也可能比2小。

如果我们定义一个新对象来表示卡牌，则其属性显然应该是rank（大小）和suit（花色）。但属性的值就不那么直观了。一种可能是使用字符串，比如用'Spade'表示花色，用'Queen'表示大小。这种实现的问题之一是比较大小和花色的高低时会比较困难。

另一种方案是使用整数来给大小和花色编码。在这个语境中，“编码”意味着我们要定义一个数字到花色，或者数字到大小的映射。这种编码并不意味着它是秘密（那样就应该称为“加密”了）。

例如，下表展示了花色和对应的整数编码：

Spades → 3

Hearts → 2

Diamonds → 1

Clubs → 0

这个编码令我们可以很容易地比较卡牌；因为更大的花色映射到更大的数字上，我们可以直接使用编码来比较花色。

卡牌大小的映射相当明显；每个数字形式的大小映射到相应的整数上，而对于花牌：

Jack → 11

Queen → 12

King → 13

我使用“→”符号，是为了说明这些映射并不是Python程序的一部分。它们是程序设计的一部分，但并不在代码中直接表现。

Card类的定义如下：

```
class Card(object):
    """Represents a standard playing card."""

    def __init__(self, suit=0, rank=2):
        self.suit = suit
        self.rank = rank
```

和前面一样，`init`方法对每个属性定义一个可选形参。默认的卡牌是草花2。

要创建一个Card对象，使用你想要的花色和大小调用Card。

```
queen_of_diamonds = Card(1, 12)
```

## 18.2 类属性

为了能将Card对象打印成人们容易阅读的格式，我们需要将整数编码映射成对应的大小和花色。自然的做法是使用字符串列表。我们将这些列表赋到类属性上：

```
# 在Card类里:
```

```
suit_names = ['Clubs', 'Diamonds', 'Hearts', 'Spades']
rank_names = [None, 'Ace', '2', '3', '4', '5', '6', '7',
              '8', '9', '10', 'Jack', 'Queen', 'King']

def __str__(self):
    return '%s of %s' % (Card.rank_names[self.rank],
                          Card.suit_names[self.suit])
```

suit\_names和rank\_names这样的变量，定义在类之中，但在任何方法之外，我们称为类属性。因为它们是和类对象Card相关联的。

这个术语和suit与rank之类的变量相区别。那些称为实例属性，因为它们是和一个特定的实例相关联的。

两种属性都使用句点表示法访问。例如，在\_\_str\_\_中，self是一个Card对象，而self.rank是它的大小。相似地，Card是一个类对象，而Card.rank\_names是关联到这个类的一个字符串列表。

每个卡片都有它自己的suit和rank，但总共只有一个suit\_names和rank\_names。

综合起来，表达式Card.rank\_names[self.rank]意思是“使用对象self的属性rank作为索引，从类Card的列表rank\_names中选择对应的字符串”。

rank\_names的第一个元素是None，因为没有大小为0的卡牌。因为使用None占据了一个位置，我们就可以得到从下标2到字符串'2'这样整齐的映射。如果要避免这种操作，可以使用字典而不是列表。

利用现有的方法，可以创建并打印卡牌：

```
>>> card1 = Card(2, 11)
>>> print card1
```

图18-1展示了Card类对象和一个Card实例。Card是一个类对象，所以它的类型是type。card1的类型是Card。（为了节省空间，我没有画出suit\_names和rank\_names的内容。）

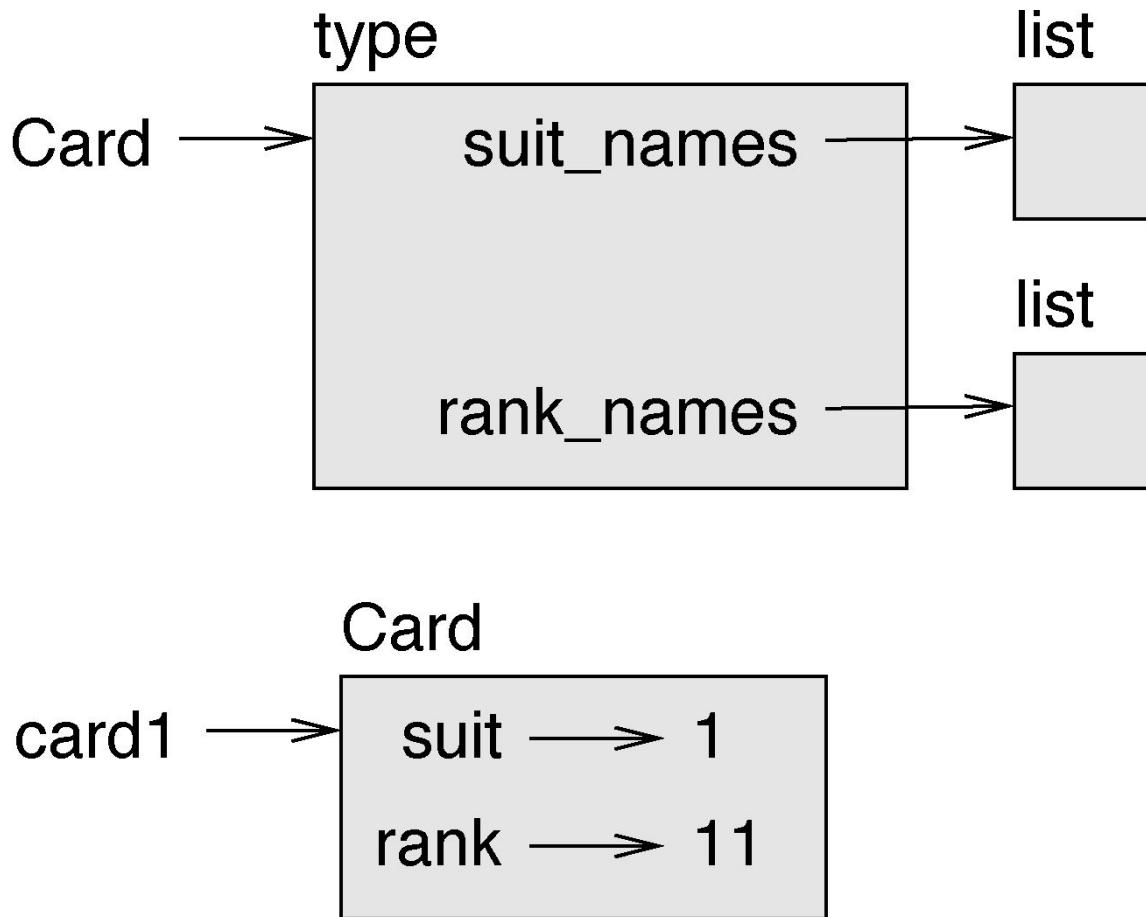


图18-1 对象图

## 18.3 对比卡牌

对于内置类型，我们用比较操作符（<、>、==等）来比较对象并决定哪一个更大、更小或者相等。对于用户定义类型，我们可以通过提供一个方法`__cmp__`来重载内置操作符的行为。

`__cmp__`接收两个形参，`self`和`other`，当第一个对象大的时候返回正数，当第二个对象大的时候返回负数，当两个对象相等时返回0。

卡牌的正确顺序并不显而易见。例如，草花3和方片2哪个更大？一个牌面数大，另一个花色大。为了比较卡牌，你需要决定大小和花色哪个更重要。

这个问题的答案取决于你在玩哪种牌类游戏，但为了简单起见，我们随意做一个决定，认为花色更重要，于是，所有的黑桃比所有的方片都大，依此类推。

这一点决定后，我们就可以编写`__cmp__`函数：

```
# 在 Card类里:

def __cmp__(self, other):
    # 检查花色
    if self.suit > other.suit: return 1
    if self.suit < other.suit: return -1

    # 花色相同，检查大小
    if self.rank > other.rank: return 1
    if self.rank < other.rank: return -1

    # 大小相同，平局
    return 0
```

使用元组比较，你可以写得更紧凑：

```
# 在Card类里:

def __cmp__(self, other):
    t1 = self.suit, self.rank
    t2 = other.suit, other.rank
    return cmp(t1, t2)
```

内置函数`cmp`和方法`__cmp__`接口一致：它接收两个值，并当第一个大的时候返回正值，当第二个大的时候返回负值，当两者相等时返回0。

### 练习18-1

为时间对象编写一个`__cmp__`方法。提示：你可以使用元组比较，但也可以考虑使用整数减法。

## 18.4 牌组

现在我们已经有了卡牌（card），下一步就是定义牌组（deck）。由于牌组是由卡牌组成的，很自然地，每个Deck对象应该有一个属性包含卡牌的列表。

下面是Deck的类定义。`init`方法创建`cards`属性，并生成52张牌的标准牌组：

```
class Deck(object):

    def __init__(self):
        self.cards = []
        for suit in range(4):
            for rank in range(1, 14):
                card = Card(suit, rank)
                self.cards.append(card)
```

填充牌组最简单的办法是使用嵌套循环。外层循环从0到3遍历各个花色。内层循环从1到13遍历卡牌大小。每次迭代使用当前的花色和大小创建一个新的Card对象，并将它添加到`self.cards`中。

## 18.5 打印牌组

下面是Deck的一个`__str__`方法：

```
# 在 Deck类里:

def __str__(self):
    res=[]
    for card in self.cards:
        res.append(str(card))
    return '\n'.join(res)
```

这个方法展示了一种累积构建大字符串的方法：先构建一个字符串的列表，再使用`join`。内置函数`str`会对每个卡牌对象调用`__str__`方法并返回字符串表达形式。

由于我们对一个换行符调用`join`函数，卡片之间用换行分隔。下面是打印的结果：

```
>>> deck = Deck()
>>> print deck
Ace of Clubs
2 of Clubs
3 of Clubs
...
10 of Spades
Jack of Spades
Queen of Spades
King of Spades
```

虽然结果显示了52行，它仍然是一个包含换行符的字符串。

## 18.6 添加、删除、洗牌和排序

为了能够发牌，我们需要一个方法从牌组中抽取一张牌并返回。列表方法`pop`为此提供了一个方便的功能：

```
# 在 Deck类里:

def pop_card(self):
    return self.cards.pop()
```

由于`pop`从列表中抽出最后一张牌，我们其实是从牌组的底端发牌的。在真实生活中“从牌组底端发牌”是会被投诉的，但在这个环境中没问题。

要添加一个卡牌，我们可以使用列表方法`append`：

```
# 在 Deck类里:

def add_card(self, card):
    self.cards.append(card)
```

像这样调用另一个函数，却不做其他更多工作的方法，有时候称为一个饰面（veneer）。这个比喻来自于木工行业。在那里，用薄薄的一层优质木料粘贴到便宜的木料表面，是一种常见工艺。

在这个例子里，我们定义了一个“浅层”方法，用更适合牌组的术语来表达一个列表操作。

作为另一个示例，我们可以使用`random`模块的函数`shuffle`来编写一个`Deck`方法`shuffle`（洗牌）：

```
# 在 Deck类里:

def shuffle(self):
    random.shuffle(self.cards)
```

不要忘记导入`random`模块。

练习18-2

编写一个Deck方法sort，使用列表方法sort来对一个Deck中的卡牌进行排序。sort使用我们定义的\_\_cmp\_\_方法来决定排序的顺序。

## 18.7 继承

和面向对象编程最相关的语言特性是继承（inheritance）。继承是一种能够定义一个新类对现有的某个类稍作修改的语言特性。

之所以被称为“继承”，是因为新类继承了现有类中的方法。继续扩展这个比喻，可以把现有类称为父类，把新类称为子类。

作为示例，假设我们想要一个类来表达一副“手牌”，即玩家手握的一副牌。一副手牌和一套牌组相似：都是由卡牌的集合组成，并且都需要诸如增加和移除卡牌的操作。

一副手牌和一套牌组也有区别；我们期望手牌拥有的一些操作，对牌组来说并无意义。例如，在扑克牌中，我们可能想要比较两副手牌来判断谁获胜了。在桥牌中，我们可能需要为一副手牌计算分数以叫牌。

这种类之间的关系——相似，但不相同——让它称为继承。

子类的定义和其他类定义相似，但父类的名称会出现在括号中：

```
class Hand(Deck):
    """Represents a hand of playing cards."""

```

这个定义说明Hand从Deck继承而来；这意味着我们可以像Deck对象那样在Hand对象上使用`pop_card`和`add_card`方法。

Hand也会继承Deck的`__init__`方法，但它和我们想要的并不一样：我们不需要填充52张卡牌，Hand的`__init__`方法应当初始化`cards`为一个空列表。

如果我们为Hand类提供一个`__init__`方法，它会覆盖Deck类的方法：

```
# 在Hand类里:
def __init__(self, label=''):
    self.cards = []
    self.label = label

```

所以当你创建Hand对象时，Python会调用这个`__init__`方法：

```
>>> hand = Hand('new hand')
>>> print hand.cards
[]
>>> print hand.label
new hand
```

但其他的方法是从Deck中继承而来的，所以我们可以使用pop\_card和add\_card来出牌：

```
>>> deck = Deck()
>>> card = deck.pop_card()
>>> hand.add_card(card)
>>> print hand
King of Spades
```

下一步很自然地就是将这段代码封装起来成为一个方法move\_cards：

```
# 在 Deck类里:

def move_cards(self, hand, num):
    for i in range(num):
        hand.add_card(self.pop_card())
```

move\_cards接收两个参数，一个Hand对象以及需要出牌的牌数。它会修改self和hand，返回None。

有的情况下，卡牌会从一副手牌中移除转入到另一幅手牌中，或者从手牌中回到牌组。你可以使用move\_cards来处理全部这些操作：self既可以是一个Deck对象，也可以是一个Hand对象。而hand参数，虽然名字是hand，却也可以是一个Deck对象。

### 练习18-3

编写一个Deck方法deal\_hands，接收两个形参：手牌的数量以及每副手牌的牌数。它会根据形参创建新的Hand对象，按照每副手牌的牌数出牌，并返回一个Hand对象列表。

继承很有用的语言特性。有些程序不用继承写，会有很多重复代码，使用继承后就会更加优雅。继承也能促进代码复用，因为可以在不修改父类的前提下对它的行为进行定制化。有的情况下，继承结构反映了问题的自然结构，所以也让程序更容易理解。

另一方面，继承也可能会让代码更难读。有时候当一个方法被调用时，并不清楚到哪里能找到它的定义。相关的代码可能分布在几个不同的模块中。并且，很多可以用继承实现的功能，也能不用它实现，甚至可以实现得更好。

## 18.8 类图

至此我们已见过用于显示程序状态的栈图，以及用于显示对象的属性和属性值的对象图。这些图表展示了程序运行中的一个快照，所以当程序继续运行时它们会跟着改变。

它们也极其详细；在某些情况下，是过于详细了。而类图对程序结构的展示相对来说更加抽象。它不会具体显示每个对象，而是显示各个类以及它们之间的关联。

类之间有下面几种关联。

- 一个类的对象可能包含其他类的对象的引用。例如，每个Rectangle对象都包含一个到Point对象的引用，而每一个Deck对象包含很多Card对象的引用。这种关联称为HAS-A（有一个），也就是说，“矩形（Rectangle）中有一个点（Point）”。
- 一个类可能继承自另一个类。这种关系称为IS-A（是一个），也就是说，“一副手牌（Hand）是一个牌组（Deck）”。
- 一个类可能依赖于另一个类，也就是说，一个类的修改需要另一个类也进行修改。

类图用图形展示了这些关系。例如，图18-2展示了Card、Deck和Hand之间的关系。

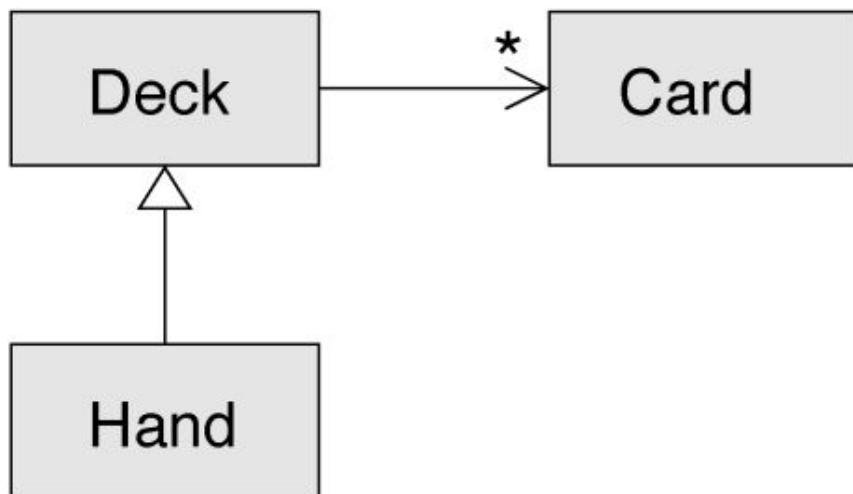


图18-2 类图

空心三角形箭头的线代表着一个IS-A关系；这里表示Hand是继承自Deck的。

标准的箭头表示HAS-A关系；这里表示Deck对象中有到Card对象的引用。

箭头附近的星号（\*）表示是关联重数标记；它表示Deck中有多少Cards。这个数可以是一个简单的数字，如52，或者一个范围，如5..7，或者一个星号，表示Deck可以有任意数量的Card引用。

更详细的图可能会显示出Deck对象实际上包含了一个Card的列表。但在类图中，像列表、字典这样的内置类型通常是不显示的。

#### 练习18-4

阅读TurtleWorld.py、World.py和Gui.py并绘制一个类图来展示其中定义的类之间的关联。

## 18.9 调试

继承会给调试带来新的挑战，因为当你调用对象的方法时，可能无法知道调用的到底是哪个方法。

假设你在编写一个操作Hand对象的函数。你可能希望能够处理所有类型的Hand，比如PokerHands、BridgeHands等。如果你调用一个方法，比如shuffle，可能调用的是Deck中定义的方法，但如果任何子类重载了这个方法，则你调用的会是那一个重载的版本。

一旦你无法确认程序的运行流程，最简单的解决办法是在相关的方法开头添加一个打印语句。如果Deck.shuffle打印一句Running Deck.shuffle这样的信息，则当程序运行时，会记录下运行的流程。

或者，你也可以使用下面这个函数。它接收一个对象和一个方法名（字符串形式），并返回提供这个方法的定义的类：

```
def find_defining_class(obj, meth_name):
    for ty in type(obj).mro():
        if meth_name in ty.__dict__:
            return ty
```

下面是使用的示例：

```
>>> hand = Hand()
>>> print find_defining_class(hand, 'shuffle')
<class 'Card.Deck'>
```

所以这个Hand对象的shuffle方法是在Deck类中定义的那个。

find\_defining\_class使用mro方法来获得用于搜索调用方法的类对象（类型）列表。“MRO”意思是“method resolution order”（方法查找顺序）。

一个设计建议：每当你重载一个方法时，新方法的接口应当和旧方法的一致。它应当接收相同的参数，返回相同的类型，并服从同样的前置条件与后置条件。如果你遵循这个规则，你会发现任何为如Deck这样的父类而设计的函数，都可以使用Hand或PokerHand这样的子类的实例。

如果你破坏这个规则，你的代码可能会像一堆纸牌屋一样崩塌（不好意思）。

## 18.10 数据封装

第16章展示了一个我们可以称为“面向对象设计”的开发计划。我们发现需要的对象——Time、Point和Rectangle——并定义类来表达它们。每个类都是一个对象到现实世界（或者最少是数学世界）中的某种实体的明显对应。

但有时候你到底需要哪些对象、它们如何交互，并不那么显而易见。这时候你需要另一种开发计划。和之前我们通过封装和泛化来发现函数接口的方式相同，我们可以通过数据封装来发现类的接口。

13.8节提供了一个很好的示例。如果你从<http://thinkpython.com/code/markov.py>下载我的代码，会发现它使用了两个全局变量——suffix\_map和prefix——并且在多个函数中进行读写。

```
suffix_map = {}
prefix = ()
```

因为这些变量是全局的，我们每次只能运行一个分析。如果我们读入两个文本，它们的前缀和后缀就会添加到相同的数据结构中（最后可以用来产生一些有趣的文本）。

若要多次运行分析，并保证它们之间的独立，我们可以将每次分析的状态信息封装成一个对象。下面是它的样子：

```
class Markov(object):

    def __init__(self):
        self.suffix_map = {}
        self.prefix = ()
```

接下来，我们将那些函数转换为方法。例如，下面是process\_word：

```
def process_word(self, word, order=2):
    if len(self.prefix) < order:
        self.prefix += (word,)
        return

    try:
```

```
    self.suffix_map[self.prefix].append(word)
except KeyError:
    # 如果这个前缀不存在, 创建一项
    self.suffix_map[self.prefix] = [word]

self.prefix = shift(self.prefix, word)
```

像这样转换程序——修改设计但不修改功能——是重构（参见4.7节）的另一个示例。

这个例子给出了一个设计对象和方法的开发计划：

1. 从编写函数、（如果需要的话）读写全局变量开始。
2. 一旦你的程序能够正确运行，查看全局变量与使用它们的函数的关联。
3. 将相关的变量封装成为对象的属性。
4. 将相关的函数转换为这个新类的方法。

### 练习18-5

从13.8节 (<http://thinkpython.com/code/markov.py>) 下载我的代码，并按照上面描述的步骤将全局变量封装为一个叫做Markov的新类的属性。解答：<http://thinkpython.com/code/ Markov.py>（注意M是大写的）。

## 18.11 术语表

编码（encode）：使用一个集合的值来表示另一个集合的值，需要在它们之间建立映射。

类属性（class attribute）：关联到类对象上的属性。类属性定义在类定义之中，但在所有方法定义之外。

实例属性（instance attribute）：和类的实例关联的属性。

饰面（veneer）：一个方法或函数，它调用另一个函数，却不做其他计算，只是为了提供不同的接口。

继承（inheritance）：可以定义一个新类，它是一个现有的类的修改版本。

父类（parent class）：被子类所继承的类。

子类（child class）：通过继承一个现有的类来创建的新类；也叫做“subclass”。

IS-A关联（IS-A relationship）：子类与父类之间的关联。

HAS-A关联（HAS-A relationship）：两个类之间的一种关联：一个类包含另一个类的对象的引用。

类图（class diagram）：用来展示程序中的类以及它们之间的关联的图。

重数（multiplicity）：类图中的一种标记方法，对于HAS-A关联，用来表示一个类中有多少对另一个类的对象的引用。

## 18.12 练习

### 练习18-6

下面列出的是扑克牌中可能的手牌，按照牌值大小的增序（也是可能性的降序）排列。

- 对子 (pair)：两张牌大小相同。
- 两对 (two pair)：两个对子。
- 三条 (three of a kind)：三张牌大小相同。
- 顺子 (straight)：五张大小相连的牌 (Ace既可以是最大也可以是最小，所以Ace-2-3-4-5是顺子，10-Jack-Queen-King-Ace也是，但Queen-King-Ace-2-3不是)。
- 同花 (flush)：五张牌花色相同。
- 满堂红 (full house)：三张牌大小相同，另外两张牌大小相同。
- 四条 (four of a kind)：四张牌大小相同。
- 同花顺 (straight flush)：顺子 (如上面的定义) 里的五张牌都是花色相同的。

本练习的目标是预测这些手牌的出牌几率。

1. 从<http://thinkpython.com/code>下载这些文件。
  - Card.py：本章中介绍的Card、Deck和Hand类的完整代码。
  - PokerHand.py：表达扑克手牌的一个类，实现并不完整，包含一些测试它的代码。
2. 如果你运行PokerHand.py，它会连出7组包含7张卡片的扑克手牌，并检查其中有没有顺子。在继续之前请仔细阅读代码。
3. 在PokerHand.py中添加方法has\_pair、has\_twopair等。它们根据手牌是否达到相对应的条件来返回True或False。你的代码应当对任意数量的手牌都适用（虽然最常见的手牌数是5或7）。
4. 编写一个函数classify（分类），它可以弄清楚一副手牌中出现的最大的组合，并设置label属性。例如，一副7张牌的手牌可能包含一个顺子以及一个对子；它应当标记为“flush”（顺子）。

5. 当你确保分类方法可用时，下一步是预测各种手牌的几率。在 `PokerHand.py` 中编写一个函数，对一副牌进行洗牌，将其分成不同手牌，对手牌进行分类，并记录每种分类出现的次数。
  6. 打印一个表格，展示各种分类以及它们的概率。更多次地运行你的程序，直到输出收敛到一个合理程度的正确性为止。将你的结果和 [http://en.wikipedia.org/wiki/Hand\\_rankings](http://en.wikipedia.org/wiki/Hand_rankings) 上的值进行对比。

解答：<http://thinkpython.com/code/PokerHandSoln.py>。
- 练习18-7
- 这个练习使用第4章中的TurtleWorld。你将会编写程序让Turtle们玩捉人游戏。如果你并不熟悉捉人游戏的规则，请参见 [http://en.wikipedia.org/wiki/Tag\\_\(game\)](http://en.wikipedia.org/wiki/Tag_(game))。
1. 下载<http://thinkpython.com/code/Wobbler.py>并运行它。你会见到一个TurtleWorld，内含3个Turtle。如果你按下运行按钮，Turtle们会随机走动。
  2. 阅读代码并保证你明白它是如何工作的。`Wobbler`类继承自 `Turtle`类，也就是说 `Turtle`类中的方法 `lt`、`rt`、`fd` 和 `bk` 都可以在 `Wobbler`类中使用。

`step`方法被 `TurtleWorld` 调用。它也调用 `steer` 来将 `Turtle` 转向期望的方向，也可以调用 `wobble` 来根据 `Turtle` 的笨拙程度随机转一个弯，以及调用 `move` 来按照 `Turtle` 的速度向前爬行几个像素。
  3. 编写一个文件 `Tagger.py`。从 `Wobbler` 中导入全部内容，并定义一个类 `Tagger` 继承 `Wobbler`。传入 `Tagger` 的类对象作为参数调用 `make_world`。
  4. 给 `Tagger` 类添加一个 `steer` 方法，覆盖 `Wobbler` 类中的方法。作为起始，编写一个版本每次都将 `Turtle` 调转朝向原点。提示：使用数学函数 `atan2` 以及 `Turtle` 的属性 `x`、`y` 和 `heading`。
  5. 修改 `steer`，让 `Turtle` 能够总保持在界内。为了调试，你可能会想要使用“单步”按钮，每按一次，对每个 `Turtle` 调用一次 `step`。
  6. 修改 `steer`，让 `Turtle` 转向最近的邻居。提示：`Turtle` 有一个属性，`world`，它是指向 `Turtle` 所生活在 `TurtleWorld` 的引用，而

TurtleWorld有一个属性，`animals`，它是世界中所有的Turtle的列表。

7. 修改`steer`，让Turtle可以玩捉人游戏。你可以给`Tagger`添加方法，并可以覆盖`steer`和`__init__`，但不可以修改或覆盖`step`、`wobble`或者`move`。另外，`steer`可以修改Turtle的指向，但不能修改它的位置。

调整规则以及你的`steer`方法来获得更优质的游戏；例如，应当让跑得慢些的Turtle能够最终追上并捉住快些的Turtle。

解答：<http://thinkpython.com/code/Tagger.py>。

# 第19章 案例研究：Tkinter

## 19.1 GUI

我们至今为止见过的大部分程序都是基于文本的，但有很多程序使用图形用户界面，也称为GUI。

Python为基于GUI的编程提供了几种选择，包括wxPython、Tkinter及Qt。每种选择都有其优劣，因此Python也没有为此形成一个标准。

本章我要展示的是Tkinter，因为我觉得它是最容易上手的。本章中介绍的大部分概念，在其他GUI框架里也适用。

关于Tkinter有几本书以及一些网页介绍。Fredrik Lundh写的*An Introduction to Tkinter*是最好的在线资源之一。

我在Swampy中编写了一个模块Gui.py。它提供了Tkinter的函数和类的简单接口。本章中的示例都是基于这个模块的。

下面是创建并显示一个GUI的简单示例。

要创建一个GUI，你需要导入Gui模块并初始化一个Gui对象：

```
from Gui import *
g = Gui()
g.title('Gui')
g.mainloop()
```

当你运行这段代码时，会出现一个灰色的方形窗口，标题是Gui。mainloop会运行事件循环（event loop），等待用户做出某些操作并进行响应。它是一个无限循环；会一直运行，直到用户关闭窗口，或者按下Control-C，或者做出某些导致程序退出的操作。

这个Gui没什么功能，因为它没有包含任何部件（widget）。部件是组成GUI的元素，它们包括以下几个。

- 按钮（button）：包含文本或图片的部件，当被按下时会进行一个操作。

- 画布 (canvas) : 一片区域，里面可以显示线段、矩形、圆圈以及其他图形。
- 输入框 (entry) : 用户可以输入文本的区域。
- 滚动条 (scrollbar) : 控制另一个部件的可见区域的部件。
- 画面框 (frame) : 包含其他部件的容器，通常不可见。

当你创建一个Gui时看到的灰色的空方块是一个画面框。当你创建一个新部件时，它会加入到这个框中。

## 19.2 按钮和回调

方法bu创建一个按钮部件：

```
button = g.bu(text='Press me.')
```

bu的返回值是一个Button对象。在框中显示的按钮就是这个对象的图形展示形式；你可以通过调用它的方法来控制这个按钮。

bu接收最多有32个形参，来控制按钮的样式与功能。这些形参称为选项（option）。你不需要提供全部32个选项的值，而是可以使用类似text='Press me.'这样的键值实参来指定你需要的选项，其他的选项会使用默认值。

当你向画面框添加一个部件时，它会“压缩”；也就是说，画面框会缩略成按钮的尺寸。如果你添加更多的部件，画面框会随之放大来适应它们的尺寸。

方法la创建一个标签（Label）部件：

```
label = g.la(text='Press the button.')
```

默认情况下，Tkinter从上到下排列控件，并且将它们居中安放。后面我们会看到如何重设这种行为。

如果你按下按钮，会发现它什么也没做。这是因为你还没有把它“组装起来”；也就是说，还没有告诉它要做什么！

用来控制按钮行为的选项是command。command的值是一个函数，当按钮被按下时会被调用。例如，下面是创建一个新标签部件的函数：

```
def make_label():
    g.la(text='Thank you.')
```

现在我们可以创建一个按钮，将它的command设置为这个函数：

```
button2 = g.bu(text='No, press me!', command=make_label)
```

当你按下这个按钮时，它会执行`make_label`函数，这时会显示出一个新的标签部件。

`command`选项的值是一个函数对象，它通常被称为回调，因为在你创建好按钮之后，当用户按下按钮时，程序执行的流程会“回调”到这个函数。

这种流程是事件驱动编程的典型特征。用户的行为，比如按下按钮或者敲击键盘等，称为事件。在事件驱动编程中，程序的流程取决于用户的行为，而不是程序员。

事件驱动编程的挑战是要构建一系列部件和回调，并保证它们在任何用户操作下都能正确工作（或者最少能生成合适的错误信息）。

### 练习19-1

编写一个程序，创建一个GUI，包含一个按钮。当按钮被按下时，它应当创建另一个按钮。当那另一个按钮被按下时，它应当创建一个写着“做得不错！”的标签部件。

当你多次按下按钮时会发生什么？

解答：[http://thinkpython.com/code/button\\_demo.py](http://thinkpython.com/code/button_demo.py)。

## 19.3 画布部件

画布是功能最丰富的部件之一。它创建一个区域，可以用来绘制线条、圆圈以及其他形状。如果你做过练习15-4，应当已经对画布很熟悉了。

方法ca创建一个新的Canvas对象：

```
canvas = g.ca(width=500, height=500)
```

width和height是画布的尺寸，单位是像素。

创建好一个部件后，可以使用config方法再次修改各个选项值。例如，bg选项修改背景颜色：

```
canvas.config(bg='white')
```

bg的值是一个表示颜色的字符串。这个值的合法集合在不同的Python实现中并不一样，但所有的实现至少都提供了下面这些值：

white	black	
red	green	blue
cyan	yellow	magenta

画布中的形状称为图形项（item）。例如，Canvas方法circle绘制（你应该猜到了）一个圆圈：

```
item = canvas.circle([0,0], 100, fill='red')
```

第一个形参是个坐标对，它指定圆心的位置；第二个形参是半径。

Gui.py提供了标准的笛卡儿坐标系，原点在画布的中心，y轴的正向指向上方。这和其他的图形系统有所区别，它们的原点在左上角，y轴的正向指向下方。

fill选项指定圆圈应当用红色填充。

circle方法的返回值是一个Item对象，它提供了一些方法用来修改画布中的图形项。例如，你可以使用config来修改圆圈的任何选项：

```
item.config(fill='yellow', outline='orange', width=10)
```

width是边界的厚度，单位是像素；outline是边界的颜色。

### 练习19-2

编写一段程序，创建一个画布和一个按钮。当用户按下按钮时，它应当在画布上绘制一个圆圈。

## 19.4 坐标序列

`rectangle`方法接收一坐标序列用来指定矩形对角的顶点。下面的示例绘制一个绿色的矩形，其左下角是原点，右上角是(200, 100)：

```
canvas.rectangle([[0, 0], [200, 100]],  
                fill='blue', outline='orange', width=10)
```

这种指定对角顶点的方式称为边界盒，因为两个点限定了矩形的边界。

`oval`接收一个边界盒，并在指定的矩形中绘制一个椭圆：

```
canvas.oval([[0, 0], [200, 100]], outline='orange', width=10)
```

`line`接收一坐标序列，并绘制一条折线连接各个点。下面的示例绘制一个三角形的两条边线：

```
canvas.line([[0, 100], [100, 200], [200, 100]], width=10)
```

`polygon`接收相同的形参，但它会（在需要时）绘制多边形的最后一边，并填充颜色：

```
canvas.polygon([[0, 100], [100, 200], [200, 100]],  
               fill='red', outline='orange', width=10)
```

## 19.5 更多部件

Tkinter提供了两个部件可以让用户输入文本：一个输入框部件（Entry），它是单行的；以及一个文本框部件（Text），它有多行。

en新建一个输入框：

```
entry = g.en(text='Default text.')
```

text选项允许你指定输入框在创建后显示的文本。get方法返回输入框的内容（可能已经被用户改变了）：

```
>>> entry.get()  
'Default text.'
```

te创建一个文本框部件：

```
text = g.te(width=100, height=5)
```

width和height是部件的尺寸，单位分别是字符数和行数。

insert可以往文本框部件中插入文本：

```
text.insert(END, 'A line of text.')
```

END是一个特殊索引，表示文本框部件的最后一个字符。

你也可以使用一个点式索引指定输入的位置，例如1.1，句点之前是行号，句点之后是列号。下面的示例在第一行的第一个字符后面添加字母'nother'。

```
>>> text.insert(1.1, 'nother')
```

get方法读入部件中的文本；它接收开始和结尾两个索引作为形参，下面的示例返回部件的全部文本，包含换行符：

```
>>> text.get(0.0, END)  
'Another line of text.\n'
```

`delete`方法从部件中删除文本；下面的示例将除了前两个字符之外的全部文本都删除掉：

```
>>> text.delete(1.2, END)
>>> text.get(0.0, END)
'An\n'
```

### 练习19-3

修改你对练习19-2的解答，添加一个输入框部件和另一个按钮。当用户按下第二个按钮时，它应当从输入框中读取一个颜色名称，并使用那个颜色来填充圆圈。使用`config`来修改已有的圆圈；不要创建新的。

你的程序应当能够处理下列情况：用户可能尝试修改一个未创建的圆圈的颜色；颜色的名称可能是不合法的。

可以在[http://thinkpython.com/code/circle\\_demo.py](http://thinkpython.com/code/circle_demo.py)看到我的解答。

## 19.6 包装部件

至今为止我们一直将部件堆放在一列，但大部分的GUI程序布局都比这个更加复杂。例如，图 19-1展示了TurtleWorld的一个简化版本（参见第4章）。

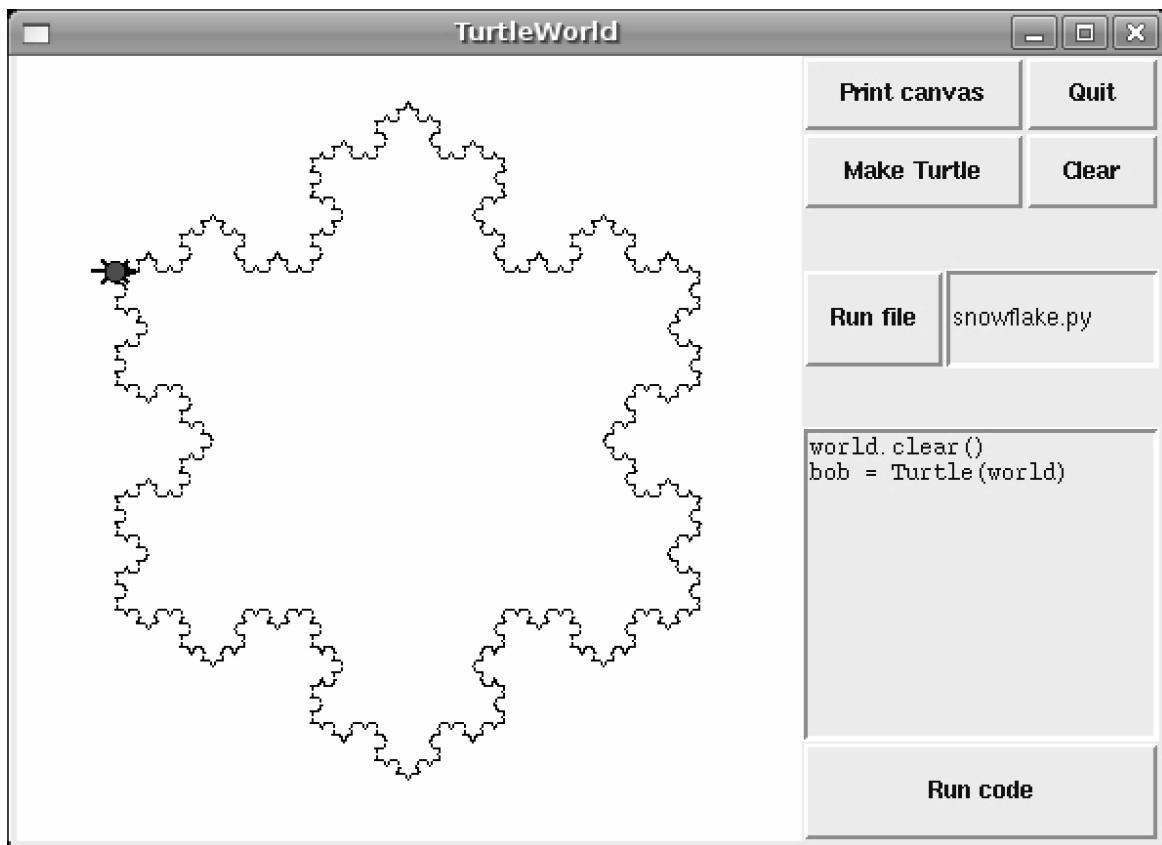


图19-1 类图

本节中分几步展示创建这个GUI的代码。你可以从  
<http://thinkpython.com/code/ SimpleTurtleWorld.py>下载完整示例的代码。

这个GUI在最顶层有了两个部件——一个画布和一个画面框——两个部件横向排布在一行中。所以第一步是创建这一行：

```
class SimpleTurtleWorld(TurtleWorld):
    """This class is identical to TurtleWorld, but the code that
    lays out the GUI is simplified for explanatory purposes."""
```

```
def setup(self):
    self.row()
    ...
```

setup是用来创建和排布部件的函数。在GUI程序中排布部件称为包装(packing)。

row创建一个单行画面框并将它设成“当前画面框”。直到这个画面框被关闭或者另一个画面框被创建之前，所有的部件都包装在这一行里。

下面的代码创建画布部件以及用来包含其他部件的列画面框：

```
self.canvas = self.ca(width=400, height=400, bg='white')
self.col()
```

列画面框中的第一个部件是一个网格画面框，内部包含4个按钮，两个两个排列：

```
self.gr(cols=2)
self.bu(text='Print canvas', command=self.canvas.dump)
self.bu(text='Quit', command=self.quit)
self.bu(text='Make Turtle', command=self.make_turtle)
self.bu(text='Clear', command=self.clear)
self.endgr()
```

gr函数创建这个网格；形参是列数。网格中的部件按照从左到右、从上到下的顺序安放。

第一个按钮使用self.canvas.dump作为回调函数；第二个使用self.quit。这些都是绑定方法，也就是说，它们都和一个特定对象关联。当它们被调用时，是在那个对象上调用的。

本列中的下一个部件是一个行画面框，包含一个按钮和一个输入框：

```
self.row([0,1], pady=30)
self.bu(text='Run file', command=self.run_file)
self.en_file = self.en(text='snowflake.py', width=5)
self.endrow()
```

row的第一个参数是一个权重列表，决定部件之间的额外空白如何分配。这里的[0, 1]表示所有的额外空白都分配给第二个部件，即文本框。当你运行这段代码并缩放窗口时，你会发现文本框会随着窗口变大，而按钮不会。

选项pady在y方向上填补这一行，在上方和下方添加30像素的空白。

endrow结束这一行部件，所以接下来的部件都会包装到列画面框中。Gui.py会保持记录画面框的栈。

- 当你使用row、col或者gr来创建一个画面框时，它会进入到栈顶并成为当前画面框。
- 当你使用endrow、endcol或者endgr来关闭一个画面框时，它会从栈顶移除，此时栈中的前一个画面框成为当前画面框。

方法run\_file读取输入框的内容，将它看做一个文件名，读入文件的内容并传给run\_code。self.inter是一个解释器对象，它知道怎么接收一段字符串并将其作为Python代码执行。

```
def run_file(self):  
    filename = self.en_file.get()  
    fp = open(filename)  
    source = fp.read()  
    self.inter.run_code(source, filename)
```

最后的两个部件是一个文本框部件和一个按钮：

```
self.te_code = self.te(width=25, height=10)  
self.te_code.insert(END, 'world.clear()\n')  
self.te_code.insert(END, 'bob = Turtle(world)\n')  
  
self.bu(text='Run code', command=self.run_text)
```

run\_text和run\_file类似，不同的是它直接从文本框中读入代码，而不需要读文件：

```
def run_text(self):  
    source = self.te_code.get(1.0, END)  
    self.inter.run_code(source, '<user-provided code>')
```

不幸的是，在其他语言中，以及其他Python模块中，部件排布的细节都不相同。Tkinter自己就提供了3种不同的机制来排布部件。这些机制称为几何管理器（geometry manager）。本章中展示的这个是“网格”（grid）几何管理器；另外两种是“包装”（pack）和“安放”（place）。

幸运的是，本节中大部分的概念在其他GUI模块或者其他语言中都适用。

## 19.7 菜单与Callable

菜单按钮（Menubutton）是一种看起来像按钮，但当它被按下时会弹出一个菜单的部件。当用户在菜单上选择某项后，菜单会消失。

下面这段代码创建一个用来选择颜色的菜单按钮（你可以从[http://thinkpython.com/code/ menubutton\\_demo.py](http://thinkpython.com/code/ menubutton_demo.py)下载）：

```
g = Gui()
g.la('Select a color:')
colors = ['red', 'green', 'blue']
mb = g.mb(text=colors[0])
```

mb创建菜单按钮。初始时，按钮上的文本是默认颜色的名称。下面的循环对每一个颜色创建一个菜单项：

```
for color in colors:
    g.mi(mb, text=color, command=Callable(set_color, color))
```

mi的第一个参数是这些菜单项关联的菜单按钮。

command选项是一个Callable对象，这是一个新概念。到此为止我们见过函数和绑定方法用作回调函数，如果你不需要传入任何实参到函数中，它们都可以良好工作。但若需要传入实参，你就需要构建一个Callable对象，它内部包含一个函数，比如set\_color，以及它的实参，比如color。

Callable对象用属性来存储到函数的引用以及各个实参。之后，当用户点击一个菜单项时，会回调这个函数，并传入存储的实参。

下面是set\_color的代码：

```
def set_color(color):
    mb.config(text=color)
    print color
```

当用户选择一个菜单项，set\_color会被调用，它会重设菜单按钮显示新选择的颜色。它也会打印出这个颜色；如果你尝试这个示例，可以确认set\_color是在你选择一项时调用的（而不是在你创建Callable对象时调用）。

## 19.8 绑定

绑定 (binding) 是一个部件、一个事件和一个回调函数之间的关联：当部件上有一个事件（比如按钮按下）发生时，回调函数会被调用。

很多部件都有默认绑定。例如，当你按下一个按钮时，默认绑定会改变按钮的样式让它看起来是被按下的。当你释放按钮时，默认绑定会让它恢复原先的样式，并调用command选项指定的回调函数。

你可以使用bind方法来覆盖这些默认绑定，或者添加新的绑定。例如，下面的代码给画布添加一个新的绑定（你可以从[http://thinkpython.com/code/draggable\\_demo.py](http://thinkpython.com/code/draggable_demo.py)下载本节的代码）：

```
ca.bind('<ButtonPress-1>', make_circle)
```

第一个形参是一个事件字符串；这个例子中，当用户使用鼠标左键按下按钮时会触发。其他的事件还包括ButtonMotion、ButtonRelease和Double-Button。

第二个形参是一个事件处理器。事件处理器是一个函数或者一个绑定方法，它类似于回调，但有一个重要的区别：事件处理器接收一个Event对象作为参数。下面是一个例子：

```
def make_circle(event):
    pos = ca.canvas_coords([event.x, event.y])
    item = ca.circle(pos, 5, fill='red')
```

Event对象包含了事件的类型信息以及其他类似于鼠标坐标点之类的具体信息。在这个例子里，我们需要的信息是鼠标单击的位置。这些值是“像素坐标”，由底层的图形系统定义。方法canvas\_coords将它们翻译成“画布坐标”，以便给circle这样的画布方法所用。

输入框部件常常会绑定<Return>事件，它在用户按回车或者换行键时触发。例如，下面的代码新建一个按钮与一个输入框：

```
bu = g.bu('Make text item:', make_text)
en = g.en()
en.bind('<Return>', make_text)
```

`make_text`在按钮被按下或者用户在输入框中按下回车键时被触发调用。要让这个功能正确运行，我们需要一个可以被当做命令（没有形参）或者事件处理器（接收Event对象作为形参）调用的函数：

```
def make_text(event=None):
    text = en.get()
    item = ca.text([0,0], text)
```

`make_text`获得输入框的内容，并在画布上用一个文本项展示它。

我们也可以为画布内部的项建立绑定。下面是一个`Draggable`的类定义，它是`Item`的子类，提供绑定来实现拖拽功能。

```
class Draggable(Item):

    def __init__(self, item):
        self.canvas = item.canvas
        self.tag = item.tag
        self.bind('<Button-3>', self.select)
        self.bind('<B3-Motion>', self.drag)
        self.bind('<Release-3>', self.drop)
```

这里的`init`方法接收一个`Item`作为参数。它复制`Item`的属性，并创建3个事件的绑定：按钮按下、按钮移动以及按钮释放。

事件处理器`select`保存当前事件的坐标以及当前项的原始值，并变成黄色：

```
def select(self, event):
    self.dragx = event.x
    self.dragy = event.y

    self.fill = self.cget('fill')
    self.config(fill='yellow')
```

`cget`代表“get configuration”（获得配置）；它接收一个选项的名称字符串并返回那个选项的当前值。

`drag`计算对象从起始点移动了多远，更新存储的坐标，然后移动它。

```
def drag(self, event):
    dx = event.x - self.dragx
    dy = event.y - self.dragy

    self.dragx = event.x
```

```
    self.dragy = event.y  
  
    self.move(dx, dy)
```

这个计算是按照像素坐标进行的；并不需要转换成画布坐标。

最后，`drop`恢复对象的原始颜色：

```
def drop(self, event):  
    self.config(fill=self.fill)
```

你可以使用`Draggable`类给一个已有的项添加拖拽功能。例如，下面是一个`make_circle`的修改版本，使用`circle`来创建一个`Item`，并使用`Draggable`来使它可以拖拽：

```
def make_circle(event):  
    pos = ca.canvas_coords([event.x, event.y])  
    item = ca.circle(pos, 5, fill='red')  
    item = Draggable(item)
```

这个例子展示了继承的一个优点：可以在不用修改父类定义的前提下，扩展或改变它的功能。当你要修改其行为的模块不是由你所写的时候，这一点特别有用。

## 19.9 调试

GUI编程的挑战之一是要记录哪些事情是在GUI正在构建时发生的，哪些事情是在之后用户行为的响应中发生的。

例如，当设置回调时，一个常见的错误是没有传入函数的引用，而是直接调用它：

```
def the_callback():
    print 'Called.'

g.bu(text='This is wrong!', command=the_callback())
```

如果你运行这段代码，会发现它立即就调用了`the_callback`，然后才创建按钮。当你按下按钮时，什么都不会发生，因为`the_callback`的返回值是`None`。通常你不应该在设置GUI的时候调用一个回调函数；它只应该在之后响应用户事件的时候被调用。

GUI编程的另一个挑战是你无法控制程序执行的流程。程序中的哪个部分会被执行，完全由用户的行动决定。这意味着你需要设计程序以能够正确处理任意顺序的事件。

例如，练习19-3中的GUI有两个部件：一个用来创建圆圈项，而另一个用来修改圆圈的颜色。如果用户先创建了圆圈再修改颜色，没有问题。但如果用户直接去修改一个还不存在的圆圈的颜色呢？或者多次创建圆圈？

当部件的数目增加后，考虑所有可能的事件序列就会变得更难。管理这种复杂度的办法之一是将系统的状态封装为一个对象并考虑以下问题。

- 有哪些可能的状态？在上面圆圈的例子中，我们可以考虑两个状态：用户创建第一个圆圈之前和之后。
- 在每个状态中，可能发生哪些事件？在这个例子中，用户可以按下任何一个按钮，或者退出。
- 对每个状态-事件对，期望的输出是什么？因为有两个状态和两个按钮，因此有4个状态-事件对需要考虑。
- 由什么导致一个状态到另一个状态的转换？在这个例子中，当用户创建第一个圆圈时，状态转换。

你也可能发现定义并检验一些在任何事件序列下都保持的不变量会很有用。

这种处理GUI编程的方式会帮助你编写正确的代码，而不需要花费时间测试每种顺序的用户事件！

## 19.10 术语表

GUI：一个图形用户界面（graphical user interface）。

部件（widget）：GUI的组成元素之一：按钮、菜单、文本输入框等。

选项（option）：用来控制一个部件的样式或功能的值。

键值参数（keyword argument）：在函数调用时带上形参名称的参数。

回调（callback）：和部件关联的函数，当用户进行某个动作时会被调用。

绑定方法（bound method）：和一个特定的实例关联的方法。

事件驱动编程（event-driven programming）：一种编程风格，程序执行的流程是由用户的行为决定的。

事件（event）：一个用户行为，比如单击鼠标或者按下键盘，可以导致GUI响应。

事件循环（event loop）：一个等待用户操作并响应的无限循环。

画布项（item）：画布部件中的一个图形元素。

边界盒（bounding box）：一个包含各种东西的矩形，通常使用两个对角顶点表示。

包装（pack）：排布和展示GUI的元素。

几何管理器（geometry manager）：用于排布部件的一个系统。

绑定（binding）：部件、事件和事件处理器之间的关联。当部件上发生事件时事件处理器会被调用。

## 19.11 练习

### 练习19-4

在本练习中，你将会编写一个图片浏览器。下面是一个简单的例子：

```
g = Gui()
canvas = g.ca(width=300)
photo = PhotoImage(file='danger.gif')
canvas.image([0,0], image=photo)
g.mainloop()
```

`PhotoImage`读入文件并返回一个`PhotoImage`对象，用于Tkinter显示。`canvas.image`将图片放到画布上，并指定其中心点的坐标。你也可以在标签、按钮以及其他一些部件上放置图片：

```
g.la(image=photo)
g.bu(image=photo)
```

`PhotoImage`只能处理少数几种格式，比如GIF和PPM，但我们可以使用Python Imaging Library (PIL，即Python图形库) 来读入其他类型的文件。

PIL模块的名称是`Image`，但Tkinter已经定义了一个名字相同的对象。为了避免这个冲突，可以使用`import...as`语句：

```
import Image as PIL
import ImageTk
```

第一行导入`Image`模块并给它一个局部名称`PIL`。第二行导入`ImageTk`，它可以将PIL图形转换为Tkinter的`PhotoImage`。下面是一个示例：

```
image = PIL.open('allen.png')
photo2 = ImageTk.PhotoImage(image)
g.la(image=photo2)
```

1. 从<http://thinkpython.com/code>下载`image_demo.py`、`danger.gif`和`allen.png`。运行`image_demo.py`。你可能需要先安装

PIL和ImageTk。它们可能已经存在于你的软件源中，但如果不在，可以从<http://pythonware.com/products/pil/>下载。

2. 在image\_demo.py中，修改第二个PhotoImage的名字，从photo2改为photo，并重新运行程序。你应当会看到第二个PhotoImage，而看不到第一个了。

问题在于当你重新给photo赋值时，它会覆盖第一个PhotoImage的引用，而它之后会消失。如果你将一个PhotoImage赋值给一个局部变量，会发生相同的事情；当函数运行结束时，它会跟着消失。

为了避免这个问题，你需要为每一个想要保持的PhotoImage存储一个引用。你可以使用全局变量，或者将PhotoImage们保存到一个数据结构或者一个对象的属性中。

这种行为可能会很恼人，所以我才在这里警告你（这也是为何示例中的图片写着“危险！”的原因）。

3. 从这个示例开始，编写一个程序，接收一个目录的名称，并遍历目录中所有的文件，将所有PIL识别出来的图片都显示出来。你可以使用try语句来捕获PIL无法识别的文件。

4. PIL提供了很多用于操作图形的方法。你可以在<http://pythonware.com/library/pil/handbook>里阅读相关的信息。作为一个挑战，选择里面的一些方法，并提供一个GUI，让用户可以对图片应用这些方法。

解答：<http://thinkpython.com/code/ImageBrowser.py>。

#### 练习19-5

矢量图形编辑器是可以让用户在屏幕上绘制并编辑图形的程序，它能够输出矢量图格式的文件，例如Postscript和SVG。

使用Tkinter编写一个简单的矢量图形编辑器。它应当最少能让用户绘制线条、圆圈和矩形，并且可以使用Canvas.dump来生成画布内容的Postscript描述。

作为挑战，你可以提供用户在画布中选择与缩放图形的功能。

#### 练习19-6

使用Tkinter来编写一个基本的网页浏览器。它应当有一个文本部件让用户输入URL，并有一个画布部件来展示页面的内容。

你可以使用`urllib`模块来下载文件（参见练习14-6），并且可以使用`HTMLParser`模块来解析HTML标签（参见<http://docs.python.org/lib/module-HTMLParser.html>）。

最起码，你的浏览器应当能处理纯文本和链接。作为挑战，你可以增加处理背景颜色、文本格式标签以及图片的功能。

# 附录A 调试

程序中可能出现不同类型的错误。区分它们可以帮助我们更快地查找以下错误。

- 语法错误由Python在将源代码翻译为字节码的过程中产生。它们通常表示程序中有语法错误。例如，在`def`语句的末尾漏掉冒号会产生一个有些冗余的错误信息`SyntaxError: invalid syntax`。
- 运行时错误由解释器在程序运行的过程中发现错误后产生。大部分错误信息都包含了错误发生的位置以及正在执行的函数的信息。例如：一个无限递归最终会导致运行时错误“`maximum recursion depth exceeded`”（超过最大递归深度）。
- 语义错误是程序运行中没有产生错误信息，但做的事情却正确的问题。例如：一个表达式求值的顺序和你预想的不同，因此产生了不正确的结果。

调试的第一步就是弄清楚你面对的到底是哪种类型的错误。虽然下面的几节是按照错误类型来组织的，但有些技巧能适用于多种情形。

## A. 1 语法错误

语法错误，在你弄清楚它们是什么之后，通常都很容易修正。不幸的是，错误信息往往没什么帮助。最常见的错误信息是 `SyntaxError: invalid syntax` 和 `SyntaxError: invalid token`，这两种都没多少信息量。

另一方面，信息也确实告诉你问题在程序中发生的位置。实际上，它告诉你的的是 Python 发现错误的位置，而并不一定总和错误发生的位置相同。有时候错误发生在错误信息指明的位置之前，往往是前一行。

如果你递增地构建程序，应当很清楚错误发生的位置。它常常在你最后添加的那行代码上。

如果你是从书本中复制代码，则最好先仔细比较你的代码和书中的代码。检查每一个字母。同时请记得书本也可能是错的，所以如果你看到一个像是语法错误的东西，那么它有可能就是。

下面是一些可以避免最常见的语法错误的方法。

1. 确保你没有使用 Python 关键字作为变量名称。
2. 检查在每一个复合语句的语句头结尾，都有一个冒号，包括 `for`、`while`、`if` 和 `def` 语句。
3. 确保程序中每个字符串都有前后匹配的引号。
4. 如果你有三引号（单引号或双引号字符）多行字符串，确保你正确结束了字符串。没有正确结束的字符串，会导致程序结尾处产生 `invalid token` 错误，或者它会将接下来的程序看作字符串的一部分，直到遇到下一个字符串为止。这种情况下，可能都不会产生错误信息！
5. 没有关闭的开始符号——`(`、`{` 或 `[`——会让 Python 继续解析下一行，并当做当前语句的一部分。通常来说，会在下一行立即产生一

个错误。

6. 检查在条件判断时将 ‘==’ 写成 ‘=’ 的经典错误。

7. 检查缩进，确保它们是按照设想正确排布的。Python可以处理空格和制表符，但如果你混合使用它们，则可能产生问题。避免这种问题最好的办法是使用一个懂得Python的编辑器，并由它产生一致的缩进。

如果上面的办法都没用，继续看下一节。

## 我一直进行修改，但没有什么区别

如果解释器报出一个错误而你找不到，有可能是因为解释器和你用的并不是同一套代码。检查你的编程环境，确保你正在编辑的代码和Python运行的是同一个。

如果不确定，可以尝试在程序开头加上一个明显而故意的错误。再运行一次。如果解释器并没有发现新的错误，那么说明你运行的不是新代码。

可能有以下几种原因。

- 你编辑了代码，但忘了保存更改就直接运行了。有的编程环境会帮你自动保存，有的不会。
- 你修改了文件名，但仍然在使用旧文件名运行程序。
- 你的编程环境可能没有正确配置。
- 如果你在编写一个模块，并使用`import`，请确保你的模块名称没有和Python标准模块冲突。
- 如果你在使用`import`来读入模块，请记得重载一个修改过的文件时，需要重启解释器或者使用`reload`。如果你直接重新导入这个模块，它并不会做任何事。

如果你遇到困难被卡住，并弄不清楚到底怎么回事，一个办法是重新以最简单的类似“Hello, World! ”的程序开始，并确保你能让一个已知的程序正确运行。然后逐渐添加原先程序的部分到新的程序中。

## A. 2 运行时错误

一旦你的程序已经确保语法正确，Python可以编译它，并最少可以开始运行它。这时候可能发生哪些错误？

### A. 2. 1 我的程序什么都不做

这个问题最常见的原因是你的文件包含了各种函数和类的定义，但没有实际调用任何代码来启动执行。如果你是为了导入模块使用它们提供的类和函数，那么这么做可能是故意的。

如果不是故意的，则确保你确实调用了一个函数来启动执行，或者在交互提示中执行了一个。另外可以参见A. 2. 5节。

### A. 2. 2 我的程序卡死了

如果一个程序突然停止并看起来什么事情都没做，它就“卡死了”。通常这意味着程序掉入一个死循环或者无限递归中。

- 如果怀疑一个特别的循环可能是问题所在，可以在循环开始前添加一个`print`语句，打出“进入循环”，在循环结尾处之后也添加一个，打出“退出循环”。再次运行程序。如果你看到第一个输出，而没有看到第二个，说明你确实遇到一个死循环了。参见A. 2. 3节。
- 大部分情况下，无限递归都会让程序运行一会儿，然后产生“`RuntimeError: Maximum recursion depth exceeded`”错误。如果发生这种情况，参见A. 2. 4节。如果你没有看到这个错误，但怀疑可能是递归方法或函数产生的问题，也同样可以使用A. 2. 4节中的技巧。
- 如果上面两步都没用，尝试其他循环或其他递归方法与函数。
- 如果这些都没用，说明可能是你没理解你的程序的执行流程。参见A. 2. 5节。

### A. 2. 3 无限循环

如果你觉得有一个无限循环并知道是哪个循环导致的问题，可以在循环的结尾处添加一个print语句，打印出循环条件中的变量值，以及条件的值。

例如：

```
while x > 0 and y < 0 :  
    # do something to x  
    # do something to y  
  
    print "x: ", x  
    print "y: ", y  
    print "condition: ", (x > 0 and y < 0)
```

现在当你再次运行程序时，能够看到每次循环中打印出的3行输出。最后一次循环时，条件应该变为false。如果循环一直进行，你应当可以看到x和y的值，并可能弄清楚为什么它们没有被正确更新。

## A. 2. 4 无限递归

大部分情况下，无限递归都会让程序运行一会儿，然后产生“Maximum recursion depth exceeded”的错误。

如果你怀疑一个函数或方法导致了无限递归，可以先检查来保证递归确实有一个基准情形。换句话说，应该有一个条件能导致函数或方法直接返回而不再继续递归调用。如果没有，那么你可能需要重新思考算法，并定位一个基准情形。

如果有一个基准情形，但程序似乎没有到达它，可以在函数或方法的开头加一个print语句来打印参数。现在当你重新运行程序时，会看到每次函数调用时都会打出几行输出，并能看到每次调用的参数。如果参数并没有向基准情形变化，你大概能发现为何如此。

## A. 2. 5 执行流程

如果你不确认程序中的执行流程如何走向，可以在每个函数的开头添加一个print语句，打印类似“进入函数foo”之类的输出。这里foo是函数名。

现在如果你重新运行程序，它会打印出每个函数调用的轨迹。

## A. 2. 6 当我运行程序，会得到一个异常

如果在运行时遇到一个问题，Python会打印出一个信息，包含错误的名称，程序中发生这个错误的位置，以及一个回溯。

回溯里标明了当前执行的函数，以及调用它的函数，以及调用这个调用者的函数，依此类推。换句话说，它回溯了从程序开头直到错误发生所在位置的整个调用轨迹。它也包括了每个函数所在的文件行号。

第一步是检查程序中错误发生的位置，并尝试弄清楚问题所在。下面是一些常见的运行时错误。

### *NameError*

你在试图使用一个当前环境中并不存在的变量。请记得局部变量是局部的。你不能在定义它们的函数之外使用它们。

### *TypeError*

有3种可能的原因。

- 你在尝试错误地使用一个值。例如：使用不是整数的值来索引字符串、列表或元组。
- 格式字符串中，内部的格式项和传入的参数不匹配。当格式项的数目不对或者转换的类型不对时都可能发生。
- 调用函数或方法时使用了错误数量的参数。对于方法来说，查看方法定义并检查第一个参数是否为`self`。接着查看方法调用；确保你是在正确类型的对象上调用方法，并正确提供了其他参数。

### *KeyError*

你在试图用一个字典并不包含的键来查找字典的元素。

### *AttributeError*

你在尝试访问一个并不存在的属性或方法。检查拼写！你可以使用`dir`函数来列出存在的属性。

如果`AttributeError`指明一个对象是`NoneType`，则意味着它是`None`。这种问题的一个常见原因是在函数里忘记返回；如果函数执行到结尾都没有遇到`return`语句，那么它会返回`None`。另一个常见的原因是使用了一个返回`None`的列表方法作为结果，比如`sort`。

### *IndexError*

你在访问列表、字符串或元组时使用的索引大于它的长度减一。在错误发生的前一行，添加一个`print`语句展示索引的值和数组的长度。数组长度是否正确？索引大小是否正确？

Python调试器（`pdb`）在查找异常时很有用，因为它让你可以在错误发生之前的地方查看程序的状态。你可以在  
<http://docs.python.org/lib/module-pdb.html>阅读`pdb`的相关资料。

## A. 2. 7 我添加了太多**print**语句，被输出淹没了

使用`print`语句进行调试的问题之一是你可能被太多的输出所埋没。有两种方法可以继续：简化输出，或者简化程序。

要简化输出，可以删除或注释掉没用的`print`语句，或者将它们合并起来，或者格式化输出让它们更容易看懂。

要简化程序，有几件事情可做。首先，简化程序所处理的问题。例如，如果你在搜索一个列表，就改为搜索一个很小的列表。如果程序从用户获得输入，则输入可以产生错误的最简单的输入。

其次，清理程序。删除无效代码，并重新组织代码让它尽可能更可读。例如，如果你怀疑问题出在程序的一个很深的嵌套部分中，则应当尝试重写那部分，让它的结构更简单。如果你怀疑一个很大的函数，则尝试将它拆分为多个更小的函数，并分别测试它们。

找寻最简测试用例的过程往往能带领你找到问题所在。如果发现程序在一种情况下正常工作，而在另一种情况下则不能，那这些情况本身就给你一些线索。

类似地，重写一部分代码可以帮助你找到细微的bug。如果你做出一个认为不该影响程序的改变，而它确实出问题了，这就给了具体的提示。

## A. 3 语义错误

从某种角度看，语义错误更难调试，因为解释器并不提供任何信息。只有你自己知道程序到底应该怎么做。

解决语义错误的第一步是在程序文本和你看到的程序行为之间建立一个连接。你需要对程序实际在做什么有一个假设。让这件事情很难的原因之一是计算机运行得太快。

你常常会希望程序能够减慢到人的速度，而使用调试器时你可以做到。但往程序里插入几条精确放置的print语句，比起设置调试器，插入或删除断点，并“单步”执行到程序出错的地方，往往花费的时间更少。

### A. 3. 1 我的程序运行不正确

你应该问自己如下几个问题。

- 程序中有没有地方你期望它去做而实际上没有发生的？找到运行那段功能的代码，并确保它确实如你所期望的那样运行了。
- 有没有一些不应该发生的事情？找到程序中运行了某种不该出现的功能的代码。
- 有没有一段代码产生的效果和你所期望的不一致？确保你完全明白该段代码，特别是当它牵涉到对其他Python模块的函数或方法的调用时。阅读你调用到的函数的文档。使用简单的测试用例测试它们并检查结果。

为了能够编程，你需要对程序如何工作有一个思维模型。如果编写出一段和你预料不同的代码，常常问题不是在程序本身，而是在你的思维模型上。

修正你的思维模型的最佳方法是将程序划分成不同部分（通常是函数和方法）并独立测试每一个部分。一旦找到你的模型和真实世界的偏差，就能够解决问题了。

当然，在开发程序时你应当分组件进行构建和测试。如果发现一个问题，应该只需要检查一小部分新的不确认是否正确的代码。

## A. 3. 2 我有一个巨大而复杂的表达式，而它和我预料的不同

编写复杂的表达式并没有问题，只要能保证它们还可读。但它们也会变得更难调试。将复杂的表达式拆分成一系列的赋值到临时变量的语句，常常是个好主意。

例如：

```
self.hands[i].addCard(self.hands[self.findNeighbor(i)].popCard())
```

这个表达式可以写作：

```
neighbor = self.findNeighbor(i)
pickedCard = self.hands[neighbor].popCard()
self.hands[i].addCard(pickedCard)
```

后面更清晰的版本也更加可读，因为变量名称提供了附加的文档信息，它也更加容易调试，因为你可以检查中间变量的类型，并打印它们的值。

复杂表达式的另一个问题是求值的顺序可能和你所期望的不同。例如，如果你将表达式 $x/2\pi$ 翻译成Python，可能会这么写：

```
y = x / 2 * math.pi
```

这样并不正确，因为乘法和除法有相同的优先级，并且语句求值的顺序是从左至右。所以这个表达式计算的实际上是 $x\pi/2$ 。

调试表达式的一个好办法是添加括号来显式控制求值顺序：

```
y = x / (2 * math.pi)
```

任何时候如果你不确定求值的顺序，都可以使用括号。这样不但会让程序更加正确（从按照你的设想来做的角度说），也会让其他人

更容易阅读你的代码，因为不需要去记忆优先级的规则。

### A. 3. 3 我有个函数或方法返回值和预期不同

如果你在程序中有`return`语句返回一个复杂的表达式，则没有机会在返回之前打印返回值。这时候，你也可以使用临时变量。例如，这个语句：

```
return self.hands[i].removeMatches()
```

可以写作：

```
count = self.hands[i].removeMatches()  
return count
```

现在你有机会在返回之前显示`count`的值了。

### A. 3. 4 我真的真的卡住了，我需要帮助

首先，试着离开计算机几分钟。计算机会发射辐射影响大脑，产生下列症状。

- 挫败感和愤怒感。
- 迷信的信念（“我的计算机恨我”）和神奇的想法（“程序只有在我反戴帽子时才正确运行”）。
- 随机行走编程（尝试着写下所有可能的程序，并选择运行正确的那个）。

如果你发现自己正在遭受这些症状之一，请马上站起来出去散个步。当你平静下来后，再思考程序。它在做什么？产生那种行为的可能原因有哪些？上一次程序还正确运行是什么时候，之后你做了什么？

有时候发现一个bug确实需要时间。我常常能够在远离计算机并让思维休息之后找到bug。找到错误的最佳地点有火车上、洗澡中及将要入睡之前。

## A. 3.5 不行，我真的需要帮助

这种事确实会发生。即使最好的程序员也会偶尔卡住。有时候你在一段程序上工作太久了所以反而看不到错误。一双新的眼睛正是所需的。

在叫人帮忙之前，请确保你已经准备好。你的程序应当尽量简单，而你应当使用最小的输入来复现错误。你应当在合适的地方放好了print语句（并且它们的输出应当容易理解）。你应当足够理解这个问题，因此能够简明扼要地描述它。

当你找人帮忙时，请确保给他们需要的信息。

- 如果有错误信息，它是什么，它代表了程序的哪部分？
- 在这个错误发生之前，你做的最后一件事情是什么？你写的最后一段代码是什么？失败的新测试用例是什么？
- 至今为止你做了哪些尝试，并从中得到了什么？

当你找寻bug时，思考一下如何做才能找得更快。下一次见到类似的情形时，就能够更快地找到问题了。

记住，目标不只是让程序正确运行。目标是学会如何让程序正确运行。

# 附录B 算法分析

这个附录编选自O'Reilly Media出版的Allen B. Downey的《Think Complexity》(2011)一书。当你阅读完本书之后，可能会想继续阅读那本书。

算法分析是计算机科学的一个分支，研究算法的性能，尤其是它们的运行时间和空间需求。参见  
[http://en.wikipedia.org/wiki/Analysis\\_of\\_algorithms](http://en.wikipedia.org/wiki/Analysis_of_algorithms)。

算法分析的实践目标是预测不同算法的性能，以便于指导设计决策。

在2008年的美国总统大选中，候选人巴拉克·奥巴马在访问谷歌公司时被要求做一个即兴分析。谷歌的首席执行官埃里克·施密特问他“给100万个32位整数排序的最高效算法”是什么。奥巴马显然被提示了，因为他马上回答，“我觉得冒泡排序可能是错误的做法”。参见  
[http://www.youtube.com/watch?v=k4RRi\\_ntQc8](http://www.youtube.com/watch?v=k4RRi_ntQc8)。

这是真的：冒泡排序在概念上很简单，但对于大数据量的排序很慢。施密特想得到的答案可能是“基数排序”  
([http://en.wikipedia.org/wiki/Radix\\_sort](http://en.wikipedia.org/wiki/Radix_sort))<sup>①</sup>。

算法分析的目标是在不同算法间做出有意义的比较，但也有一些问题。

- 算法的相对性能可能依赖于硬件的特征，所以一个算法可能在机器A上更快，另一个在机器B上更快。这个问题的通用解决方法是先指定一个机器模型，并分析在一个指定的机器模型中一个算法需要执行的步骤或操作。
- 相对性能还可能依赖于数据集的细节特征。例如，有的排序算法在数据已经是部分排序的情形下比其他算法更快；有的程序在这种情况下反而慢。避免这个问题的通常办法是分析最坏情况场景。有时候分析平均情况的性能也有用，但也通常会更难，因为有哪些情形可以用来“平均”往往并不明显。
- 相对性能也依赖于问题的尺寸。对小序列更快的排序算法可能对大序列就慢了。这个问题的通常解决方案是用一个问题尺寸的函数来表达运行时间（或指令数），并在问题尺寸增大时渐进地比较函数。

这种比较的好处之一是自然而然地可以将算法进行简单地分类。例如，如果我知道算法A的运行时间趋向于和输入的尺寸 $n$ 成比例，而算法B趋向于和 $n^2$ 成比例，那么我会预期对于大的 $n$ 值，算法A比算法B快。

这种分析也有需要注意的地方，后面会谈到。

## B. 1 量级

假设你需要分析两个算法，并依照输入的尺寸来表达它们的运行时间：算法A需要 $100n+1$ 步来解决尺寸为 $n$ 的问题；算法B需要 $n^2 + n + 1$ 步。

下面的表格显示了这两个算法在不同的问题尺寸下的运行时间：

输入尺寸	算法A的运行时间	算法B的运行时间
10	1 001	111
100	10 001	10 101
1 000	100 001	1 001 001
10 000	1 000 001	$>10^{10}$

在 $n=10$ 时，算法A看起来很差；它几乎需要10倍于算法B的时间。但对于 $n=100$ 来说它们就已经差不多了，而在更大的尺寸时，算法A远好于算法B。

这里根本的原因在于对很大的 $n$ 值，任何包含 $n^2$ 项的函数都会比首项是 $n$ 的函数增长快速很多。首项是一个多项式中最高次方的项。

对于算法A，首项有一个很大的系数100，因此算法B在小的 $n$ 时比算法A快。但不论系数是多少，总有一个 $n$ 值会导致 $an^2 > bn$ 。

对于非首项来说也如此。即使算法A的运行时间是 $n+1000000$ ，对于足够大的 $n$ ，仍然会比算法B快。

总的来说，我们预期有更小的首项的算法对大尺寸问题来说是更好的算法。但对于小一些的问题来说，可能存在一个交叉点，那里其他算法可能更好。交叉点的位置取决于算法的细节、输入以及硬件的条件，所以在算法分析时常常被忽略掉。但那并不意味着你可以忘记它。

如果两个算法有相同的首项，则很难说哪一个更好；同样地，答案也取决于细节条件。所以对于算法分析来说，首项相同的函数被认为是同等的，即使它们的系数不同。

量级就是各种渐进增长被认为是同等的函数的集合。例如， $2n$ 、 $100n$ 和 $n+1$ 都是一个量级，用大 $O$ 标记法写作 $O(n)$ ，并通常被称为线性的，因为

这个集合中每个函数都依据 $n$ 线性增长。

所有首项是 $n^2$ 的函数都属于 $O(n^2)$ ；它们被称为是平方的。平方是对首项 $n^2$ 的一个好听的叫法。

下面的表格显示了算法分析中大部分最常见的量级，按照更坏的程度递增：

量级	名称
$O(1)$	常量级
$O(\log_b n)$	对数级（对任意 $b$ ）
$O(n)$	线性级
$O(n \log_b n)$	$n \log n$
$O(n^2)$	平方级
$O(n^3)$	立方级
$O(c^n)$	指数级（底数 $c$ 任意）

对于对数项，底数并没有影响；修改底数相当于乘以一个常量，而那样并不影响量级。类似地，所有的指数函数都是同一个量级，不论指数的底数是什么。指数函数增长非常迅速，所以指数级算法只在小尺寸问题中应用。

### 练习B-1

在[http://en.wikipedia.org/wiki/Big\\_O\\_notation](http://en.wikipedia.org/wiki/Big_O_notation)上阅读大O标记法的维基百科页面，并回答下列问题。

1.  $n^3 + n^2$ 的量级是多少？ $1000000n^3 + n^2$ 呢？ $n^3 + 1000000n^2$ 呢？

2.  $(n^2 + n) \cdot (n + 1)$  的量级是多少？在相乘之前，请记住你只需要首项。
3. 如果  $f$  是  $\mathcal{O}(g)$ ，对于未指定的函数  $g$ ，我们怎么说  $af + b$ ？
4. 如果  $f_1$  和  $f_2$  都是  $\mathcal{O}(g)$ ，那么  $f_1 + f_2$  呢？
5. 如果  $f_1$  是  $\mathcal{O}(g)$  而  $f_2$  是  $\mathcal{O}(h)$ ，那么  $f_1 + f_2$  呢？
6. 如果  $f_1$  是  $\mathcal{O}(g)$  而  $f_2$  是  $\mathcal{O}(h)$ ，那么  $f_1 \cdot f_2$  呢？

关心程序性能的程序员常常会觉得这种分析很难理解。他们有道理：有时候系数和非首项也能带来不同。有时候硬件的细节、编程语言，以及输入的特征，都能带来很大的区别。并且对于小尺寸问题来说，渐进行为是无关要紧的。

但如果在脑中记着这些需要注意的要点的话，算法分析毕竟是一个有用的工具。至少对于大尺寸问题来说，“更好”的算法往往确实更好，并且有时候它会好得多。两个量级相同的算法的区别往往是一个常量值，但一个好算法和一个坏算法的差距是没有界限的！

## B. 2 Python基本操作的分析

大部分算术操作都是常量时间的；乘法通常比加法和减法花费更多时间，而除法花费的更多，但这些操作的时间并不依赖于参数的大小。特别大的整数是一个例外；在那种情况下，运行时间随着数字的位数增加而增加。

索引操作——在序列或字典中读写元素——也是常量时间的，与数据结构的尺寸无关。

遍历一个序列或字典的`for`循环通常是线性的，只要循环体内的操作本身是常量级。例如，将一个列表的元素相加是线性的：

```
total = 0
for x in t:
    total += x
```

内置函数`sum`也是线性的，因为它做相同的事情。但它趋向于更快些，因为实现得更高效；用算法分析的语言来说，就是它有一个更小的首项系数。

如果你使用相同的循环来“累加”一个字符串的列表，则运行时间是平方的，因为字符串拼接本身是线性的。

字符串方法`join`通常更快，因为它是按照字符串的总长度线性的。

作为一个经验规则，如果循环体的量级是 $O(n^a)$ 则整个循环是 $O(n^{a+1})$ 。例外情况是当你能够证明循环在一个常量数的迭代之后就能退出。如果不論 $n$ 是多少，循环只最多运行 $k$ 次，则即使对很大的 $k$ 来说，整个循环的量级还是 $O(n^a)$ 。

大部分字符串和元组操作都是线性的，但如果字符串的长度是有常量界限的——例如，操作在单一字符上——则可以看做是常量时间的。

大多数列表方法是线性的，但也有一些例外。

- 在列表结尾处添加一个元素的操作平均来说是常量时间的；当它空间不足时，偶尔会复制到另一个更大的地方，但总的 $n$ 次操作的时间量级

是 $\mathcal{O}(n)$ ，所以我们说一次操作的“均摊”(amortized)的时间是 $\mathcal{O}(1)$ 。

- 从列表结尾删除一个元素的操作是常量时间的。
- 排序的量级是 $\mathcal{O}(n \log n)$ 。

大部分字典操作和方法都是常量时间的，但也有一些例外。

- `copy`的运行时间和元素的数量成比例，但和元素的大小无关（它复制引用，而不是元素本身）。
- `update`的运行时间和作为参数传入的字典的大小成比例，而不是被更新的字典本身。
- `keys`、`values`和`items`都是线性的，因为它们返回新的列表；`iterkeys`、`itervalues`和`iteritems`是常量时间的，因为它们返回的是迭代器。但如果你循环遍历这个迭代器，则循环是线性的。使用“`iter`”函数节省了一些额外开支，但并没有改变操作的量级，除非你访问的元素的数量是受限制的。

字典的效率是计算机科学的一个小奇迹。我们会在B.4节中介绍它。

## 练习B-2

在[http://en.wikipedia.org/wiki/Sorting\\_algorithm](http://en.wikipedia.org/wiki/Sorting_algorithm)里阅读排序算法的维基百科页面并回答下列问题。

1. 什么是“比较排序”？比较排序的最坏情况的量级最好是什么？任何排序算法中，最坏情况的量级最好是多少？
2. 冒泡排序的量级是多少？为什么奥巴马认为它是“错误的做法”？
3. 基数排序的量级是多少？要使用它，我们需要哪些前置条件？
4. 稳定排序是什么，为什么在实践中它很重要？
5. 最差的（有名字的）排序算法是什么？
6. C语言库里用的排序算法是什么？Python里用的是什么？这些算法有文档吗？你可能需要去谷歌搜索这些答案。
7. 很多非比较排序都是线性的，那么为什么Python会使用一个 $\mathcal{O}(n \log n)$ 的比较排序呢？

## B.3 搜索算法的分析

搜索是一种算法，接收一个集合和一个目标元素，并决定这个元素是否在集合中，通常返回元素的索引。

最简单的搜索算法是“线性搜索”，即按顺序遍历集合的每一个元素，直到找到目标元素为止。在最坏的情况下，它会遍历整个集合，所以运行时间是线性的。

序列的`in`操作符使用一个线性搜索；字符串方法`find`和`count`也是这样。

如果序列中的元素是排好序的，你可以使用二分查找，它的量级是 $O(\log n)$ 。二分查找和你在字典（真实的字典，而不是那个数据结构）中查找单词的算法类似。不像普通搜索那样从第一个元素开始，它是从序列的中间开始，检查你要查找的词是在中间的元素之前还是之后。如果在之前，则继续查找序列的前半段，否则查找后半段。不论哪种情况，你都可以将查找的数量减少一半。

如果序列有1 000 000个元素，大概需要花20个步骤找到单词或者发现它不存在。所以那样会比线性查找快大概50 000倍。

### 练习B-3

编写一个函数`bisection`，接收一个排好序的列表和一个目标值，并返回目标值所在的索引，如果不在列表中，返回`None`。

或者你可以阅读`bisect`模块的文档并使用它！

二分查找可以比线性查找快很多，但需要序列本身是排好序的，也就需要一些额外工作。

有另一个数据结构，称为散列表（hashtable），它甚至更快——它可以用常量时间来搜索——而且不需要元素是排好序的。Python字典是使用散列表实现的，因此大部分字典操作，包括`in`操作符，都是常量时间的。

## B. 4 散列表

为了解释散列表的工作机制以及为何它的效率如此好，我们先从一个简单的映射实现开始，并逐步改善它，直到成为一个散列表。

我使用Python来展示这些实现。但真实世界中你不需要用Python写这样的代码；你只需要直接使用字典即可！所以本章中剩下的部分，你需要想象字典并不存在，而你需要实现一个数据结构将键映射到值。你需要实现的操作有：

```
add(k, v):
```

添加一个新项，将键k映射到值v。在Python字典d中，这个操作写作`d[k] = v`。

```
get(target):
```

根据键target查找对应的值。在Python字典d中，这个操作写作`d[target]`或`d.get(target)`。就现在来说，我假设每个键只出现一次。最简单的实现是使用一个元组列表，每个元组是一个键值对。

```
class LinearMap(object):  
    def __init__(self):  
        self.items = []  
  
    def add(self, k, v):  
        self.items.append((k, v))  
  
    def get(self, k):  
        for key, val in self.items:  
            if key == k:  
                return val  
        raise KeyError
```

`add`往元组列表中添加一项，这个操作是常量时间的。

`get`使用一个`for`循环来搜索列表：如果找到了目标键，则返回对应的值；否则抛出`KeyError`。所以`get`是线性的。

另一个方案是让列表按照键来排序。这样get就可以使用二分查找，其量级是 $O(\log n)$ 。但插入一个新项到列表中间是线性的，所以这可能也不是最好的选择。也有数据结构（参见[http://en.wikipedia.org/wiki/Red-black\\_tree](http://en.wikipedia.org/wiki/Red-black_tree)）可以用对数时间实现add和get，但那仍然没有常量时间好，所以我们继续。

优化LinearMap的方法之一是将键值对的列表拆分成更小的列表。下面是一个称为BetterMap的实现，它是一个包含100个LinearMap的列表。我们接下来会看到，get的量级仍然是线性的，但是BetterMap离散列表更近了一步。

```
class BetterMap(object):

    def __init__(self, n=100):
        self.maps = []
        for i in range(n):
            self.maps.append(LinearMap())

    def find_map(self, k):
        index = hash(k) % len(self.maps)
        return self.maps[index]

    def add(self, k, v):
        m = self.find_map(k)
        m.add(k, v)

    def get(self, k):
        m = self.find_map(k)
        return m.get(k)
```

`__init__` 创建由n个LinearMap组成的列表。

`find_map`被`add`和`get`调用，用来确定用哪个映射来保存新项，或者到哪个映射里去搜索。

`find_map`使用了内置函数`hash`，它接收几乎所有的Python对象，并返回一个整数。这个实现的限制之一是它只对可散列的键类型可用。可变类型，如列表和字典，是不可散列的。

两个认为相等的可散列对象会返回相同的散列值，但反过来并不一定是对的：两个不同的对象可以返回相同的散列值。

`find_map`使用求余操作符来将散列值封装到0到`len(self.maps)`的范围内，这样结果是列表的一个合法索引。当然，这意味着很多不同的散

列值会封装到同一个索引上。但如散列函数将对象分配地很均匀（这也是散列函数设计的目标），那么我们预计每个LinearMap有 $n/100$ 个项。

因为LinearMap.get的运行时间是和其包含的项数成比例的，所以我们预计BetterMap会比LinearMap快100倍。量级仍然是线性，但首项系数更小。这很好，但仍然不如散列表好。

下面（终于）是让散列表能变快的关键原因：如果你能保证LinearMap的长度有限，LinearMap.get则会是常量时间。你需要做的只是记录元素的总数，并当每个LinearMap的大小超过一个阈值时，重新划分散列表，添加更多的LinearMap。

下面是一个散列表的实现：

```
class HashMap(object):

    def __init__(self):
        self.maps = BetterMap(2)
        self.num = 0

    def get(self, k):
        return self.maps.get(k)

    def add(self, k, v):
        if self.num == len(self.maps.maps):
            self.resize()

        self.maps.add(k, v)
        self.num += 1

    def resize(self):
        new_maps = BetterMap(self.num * 2)

        for m in self.maps.maps:
            for k, v in m.items():
                new_maps.add(k, v)

        self.maps = new_maps
```

每个HashMap都包含一个BetterMap；\_\_init\_\_从2个LinearMap开始，并初始化num，它会用来记录总的项数。

get只需要分配到对应的BetterMap。真正的工作都发生在add中，它会检查项数和BetterMap的大小：如果相等，那么每个LinearMap的平均项数是1，所以它调用resize。

`resize`创建一个新的BetterMap，比之前大一倍，并将旧有的映射中的项“重新散列”到新的映射中。

重散列是有必要的，因为LinearMap的数量的改变，导致`find_map`的求余操作符的分母改变。也就是说，有些原先会封装到同一个LinearMap的项会分配到不同的LinearMap中（这也是我们想要的，对吧？）。

重散列是线性的，所以`resize`是线性的，看起来可能不好，因为我保证过`add`应当是常量时间的。但请记得我们并不是每次都需要进行`resize`，所以`add`通常是常量时间的，只是偶尔会线性。`add`运行 $n$ 次的总时间是和 $n$ 成比例的，因此每次调用`add`的平均时间是常量时间！

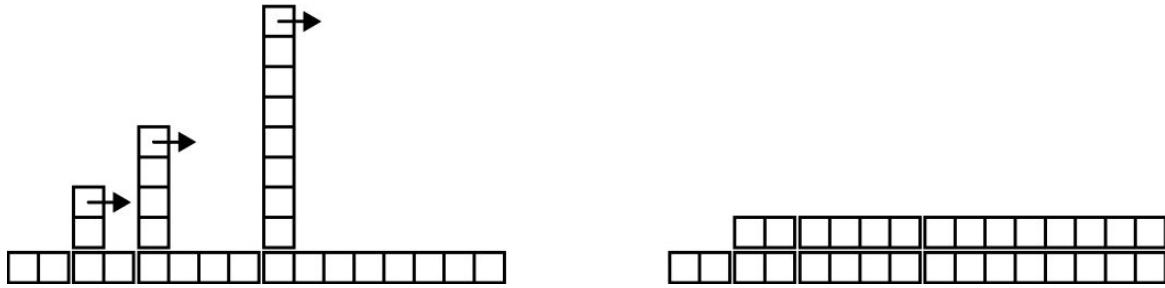
要明白散列表如何工作，考虑从一个空的HashTable开始，并添加一序列项。我们从2个LinearMap开始，所以最开始两个`add`会很快（不需要`resize`）。我们说它们每次花费一单位的工作量。下一个`add`会需要`resize`，所以我们需要重散列前两项（我们说这需要再加2个单位的工作量）并添加一个新项（再加1个单位）。再添加一项花费1单位，所以至今为止是4项花费了6个单位的工作。

下一个`add`需要5个单位，但接着的3个都只需要1个单位，所以总共是8项花费了14单位。

再下一个`add`需要9个单位，但接着我们可以在再次`resize`之前添加7项，所以总共是16个`add`花费了30单位。

在32个`add`时，总共的花费是62单位，而我希望你已经开始看到其中的模式了。在 $n$ 个`add`之后，假设 $n$ 是2的乘方，总的花费是 $2n - 2$ 单位，所以平均每个`add`的工作量是稍微小于2个单位的。当 $n$ 是2的乘方时，这是最好情况；对于其他的 $n$ 值，平均工作量稍高一点，但这并不重要。重要的是这是 $O(1)$ 。

图B-1用图形化的方式展示了这个过程。每个方块代表一个单位的工作量。每一列显示每个`add`的工作量：从左到右，前两个`add`花费1单位，第三个花费3单位，等等。



图B-1 散列表add的消耗

多余的重散列的工作看起来像一序列不断增高的塔，之间的间隔越来越远。现在如果你将塔推倒，将`resize`的花费均摊到所有`add`操作上，就会发现 $n$ 个`add`之后总的花费是 $2n-2$ 。

这个算法的一个重要特点是当我们调整`HashTable`的大小时，它会几何增长；也就是，我们乘以一个常量到大小上。如果你算术地增加大小——每次添加固定数量的数——那么每个`add`的平均时间是线性的。

你可以从<http://thinkpython.com/code/Map.py>下载我的`HashMap`实现，但请记住并没有使用它的理由；如果你需要一个映射，直接用Python字典即可。

① 但如果你在面试时被问到这个问题，我觉得更好的答案是，“给100万个数排序的最快算法应当是使用我用的语言提供的排序函数。它的性能应当对绝大多数应用都足够好了，但如果发现我的程序太慢，我会使用一个性能分析器去查看时间花在哪里。如果看起来更快的排序算法会带来明显的提升，那我会去寻找一个基数排序的良好实现。”

## 附录C Lumpy

在本书中，我使用图表来展示程序运行的状态。

在2.2节中，我们使用了一个状态图来展示变量的名称和值。在3.10节中，我介绍了栈图，用来展示每个函数调用的帧状态。递归函数的栈图出现在5.9节和6.5节中。

10.2节中使用状态图展示了列表的样子，11.4节中展示了字典的样子，以及12.6节中展示了表示元组的两种方式。

15.2节中介绍了对象图，展示一个对象的状态，包括它的属性，以及属性的属性，如此递归。15.3节中有一个对象图展示Rectangle和它嵌套的Point对象。16.1节展示了一个时间对象的状态。18.2节里有一个图，包含一个类对象和一个实例对象，各自有它们自己的属性。

最后，18.8节介绍了类图，用来展示组成程序的类和它们之间的关系。

这些图表都基于统一建模语言（UML，Unified Modeling Language），它是一个标准化的图形语言，软件工程师用这种语言来交流程序设计，特别是面向对象设计。

UML是一个丰富的语言，有很多种图表，可以表达对象和类之间的各种关系。本书中我介绍的只是它的一个很小的子集，但也是实践中最常用的子集。

本附录的目标是检视前面各章中出现的图表，并介绍Lumpy。Lumpy代表“UML in Python”，将其中的一些字母重安排后形成这个名字。Lumpy是Swampy的一部分，它在第4章或第19章中的案例分析中已有介绍，或者如果你做过练习15-4，应该已经安装过。

Lumpy使用了Python的inspect模块来检查一个运行中的程序的状态，并生成对象图（包括栈图）以及类图。

## C. 1 状态图

下面是用Lumpy生成状态图的示例：

```
from swampy.Lumpy import Lumpy

lumpy = Lumpy()
lumpy.make_reference()

message = 'And now for something completely different'
n = 17
pi = 3.1415926535897932

lumpy.object_diagram()
```

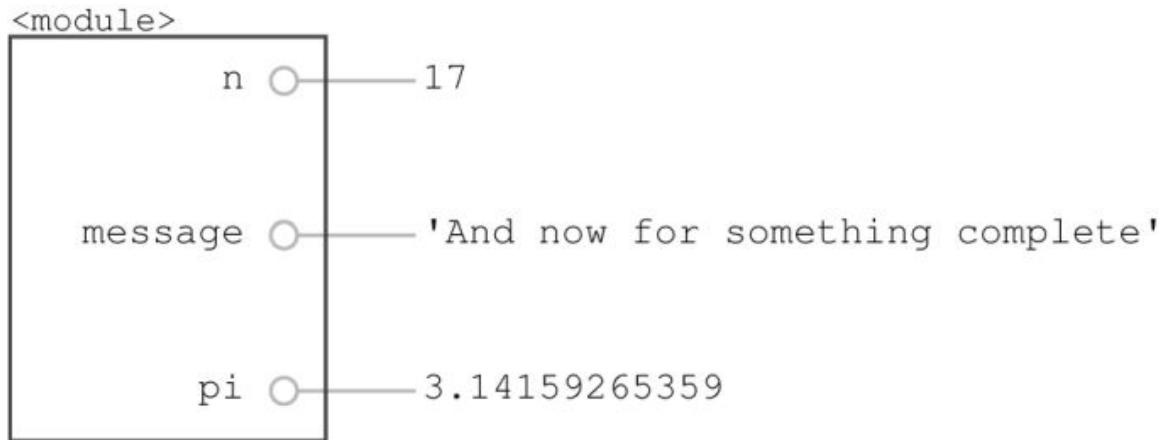
第一行从swampy.Lumpy模块中导入Lumpy类。如果你没有打包安装Swampy，请保证Swampy的文件在Python的搜索路径上，并使用下面的导入语句：

```
from Lumpy import Lumpy
```

下一行创建一个Lumpy对象，并建立一个“引用”点，也就是说Lumpy会记录到目前为止定义了的对象。

接下来我们定义新的变量并调用object\_diagram，它将绘制从引用点开始新定义的对象，在这个例子里是message、n和pi。

图C-1展示了状态图的结果。图形样式和我之前展示的不一样。例如，每个引用都是使用一个圆圈紧挨着变量名称，并用一条直线连接到引用的值。此外，很长的字符串会被截断。但是图中展示的信息是一样的。



图C-1 Lumpy生成的状态图

变量名称包括在一个标注了<module>的图框中，表示这些变量是模块级别的变量，也就是全局变量。

你可以从[http://thinkpython.com/code/lumpy\\_demo1.py](http://thinkpython.com/code/lumpy_demo1.py)下载这个示例。尝试着添加一些多余的赋值，并看看图表会如何展示。

## C. 2 栈图

下面是一个使用Lumpy生成栈图的示例。你可以从[http://thinkpython.com/code/lumpy\\_demo2.py](http://thinkpython.com/code/lumpy_demo2.py)下载它。

```
from swampy.Lumpy import Lumpy

def countdown(n):
    if n <= 0:
        print 'Blastoff!'
        lumpy.object_diagram()
    else:
        print n
        countdown(n-1)

lumpy = Lumpy()
lumpy.make_reference()
countdown(3)
```

图C-2展示栈图的结果。每一帧都使用一个盒子表示，外面是函数名称，里面是变量。因为这个函数是递归的，所以每层递归都有一帧。



图C-2 栈图

注意，栈图展示的是在程序运行中的一个特定点上的状态。要获得你需要的栈图，有时候需要思考在哪里调用`object_diagram`。

在这个例子里，我在执行完递归函数的基准情形后调用`object_diagram`来获得程序执行的一系列快照。

## C. 3 对象图

下面的示例生成10.1节中的列表的对象图。你可以从[http://thinkpython.com/code/lumpy\\_demo3.py](http://thinkpython.com/code/lumpy_demo3.py)下载它。

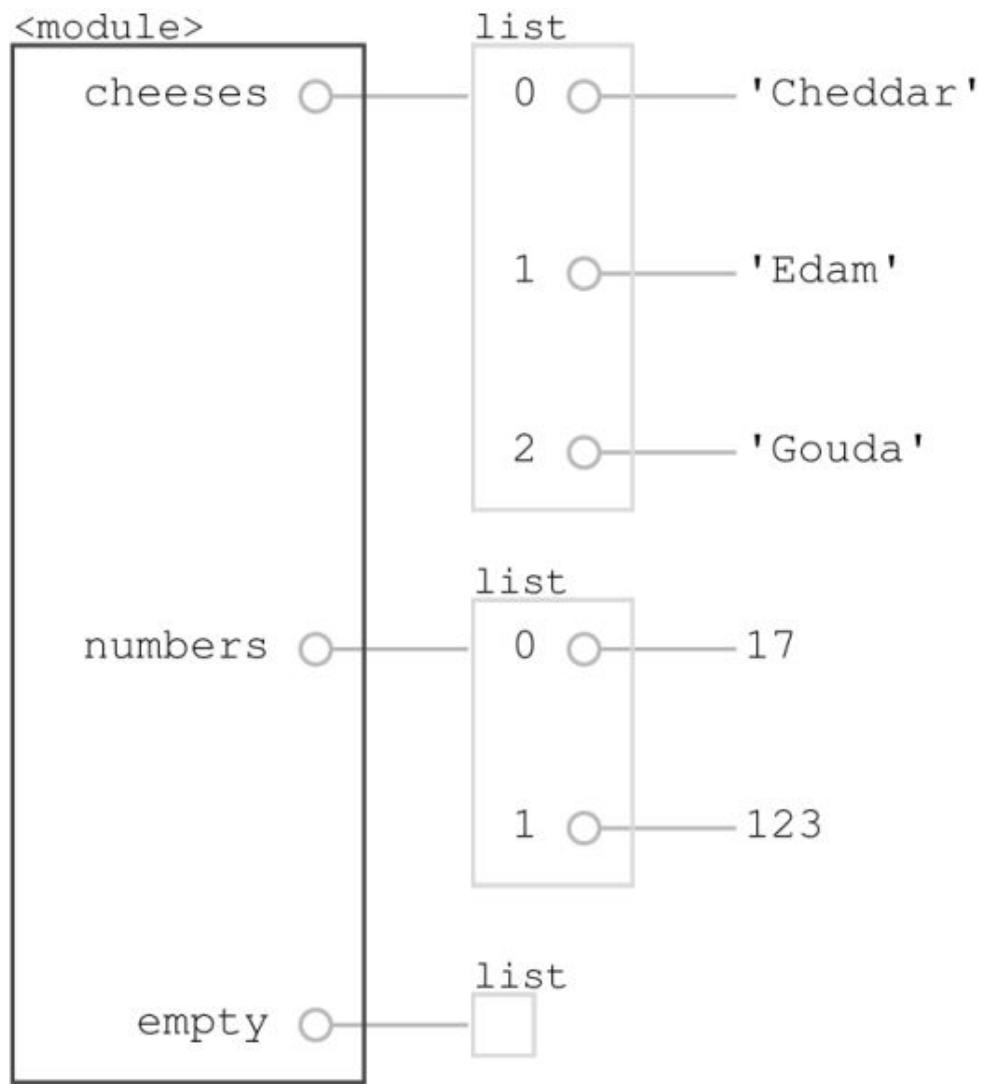
```
from swampy.Lumpy import Lumpy

lumpy = Lumpy()
lumpy.make_reference()

cheeses = ['Cheddar', 'Edam', 'Gouda']
numbers = [17, 123]
empty = []

lumpy.object_diagram()
```

图C-3展示对象图的结果。列表用一个盒子表示，里面展示了从索引到元素的映射。这个展示有些误导，因为索引实际上并不是列表的一部分，但我觉得它们让对象图更可读。空列表使用一个空盒表示。



图C-3 对象图

下面是一个展示11.4节中的字典的对象图的例子。你可以从  
[http://thinkpython.com/code/lumpy\\_demo4.py](http://thinkpython.com/code/lumpy_demo4.py)下载它。

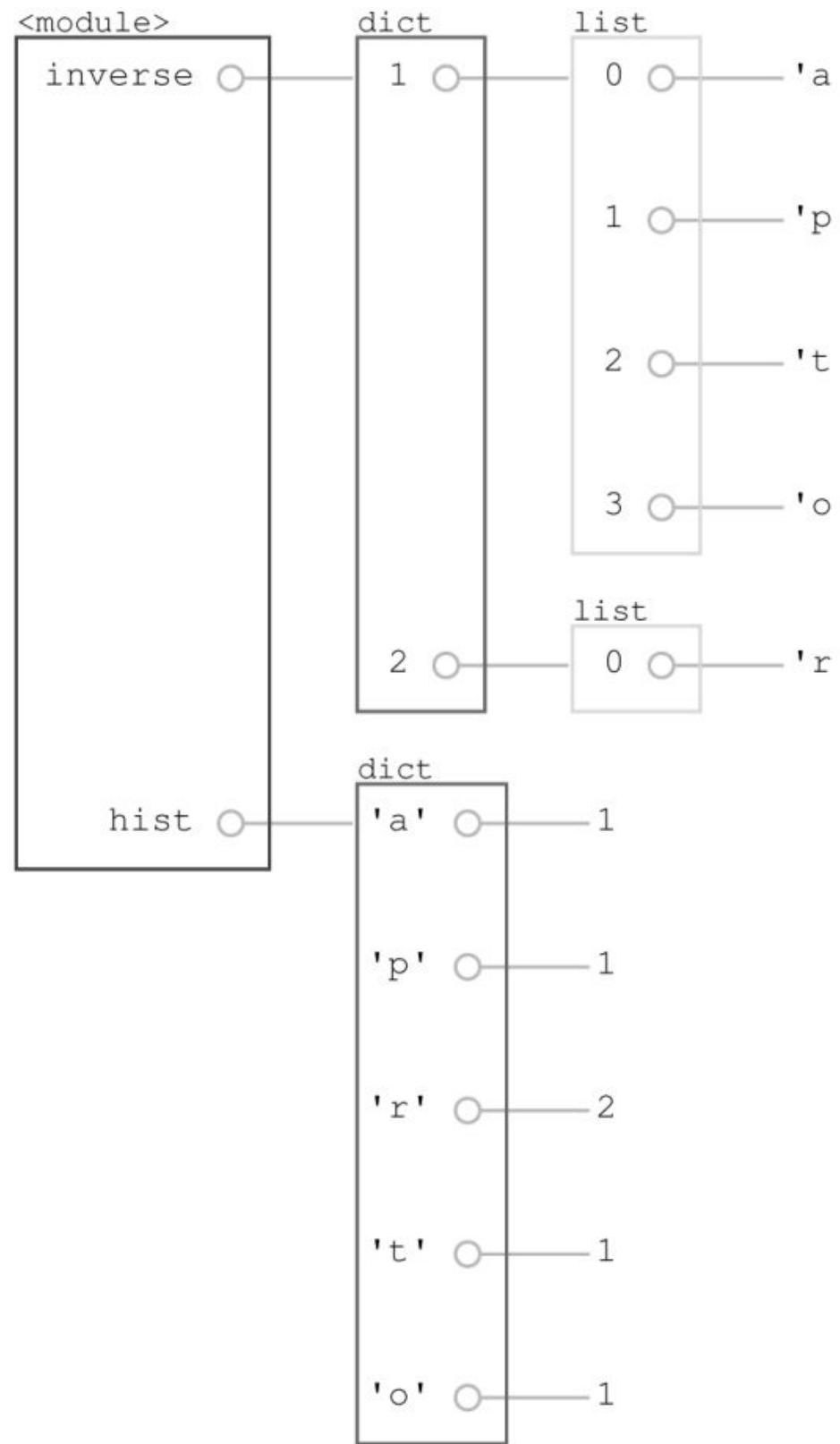
```
from swampy.Lumpy import Lumpy

lumpy = Lumpy()
lumpy.make_reference()

hist = histogram('parrot')
inverse = invert_dict(hist)

lumpy.object_diagram()
```

图C-4展示对象图的结果。hist是一个字典，将字符（单个字母的字符串）映射成整数；inverse将整数映射成字符串的列表。



图C-4 对象图

下面的示例为Point和Rectangle对象生成对象图，如15.6节中所介绍。你可以从[http://thinkpython.com/code/lumpy\\_demo5.py](http://thinkpython.com/code/lumpy_demo5.py)下载它。

```
import copy
from swampy.Lumpy import Lumpy

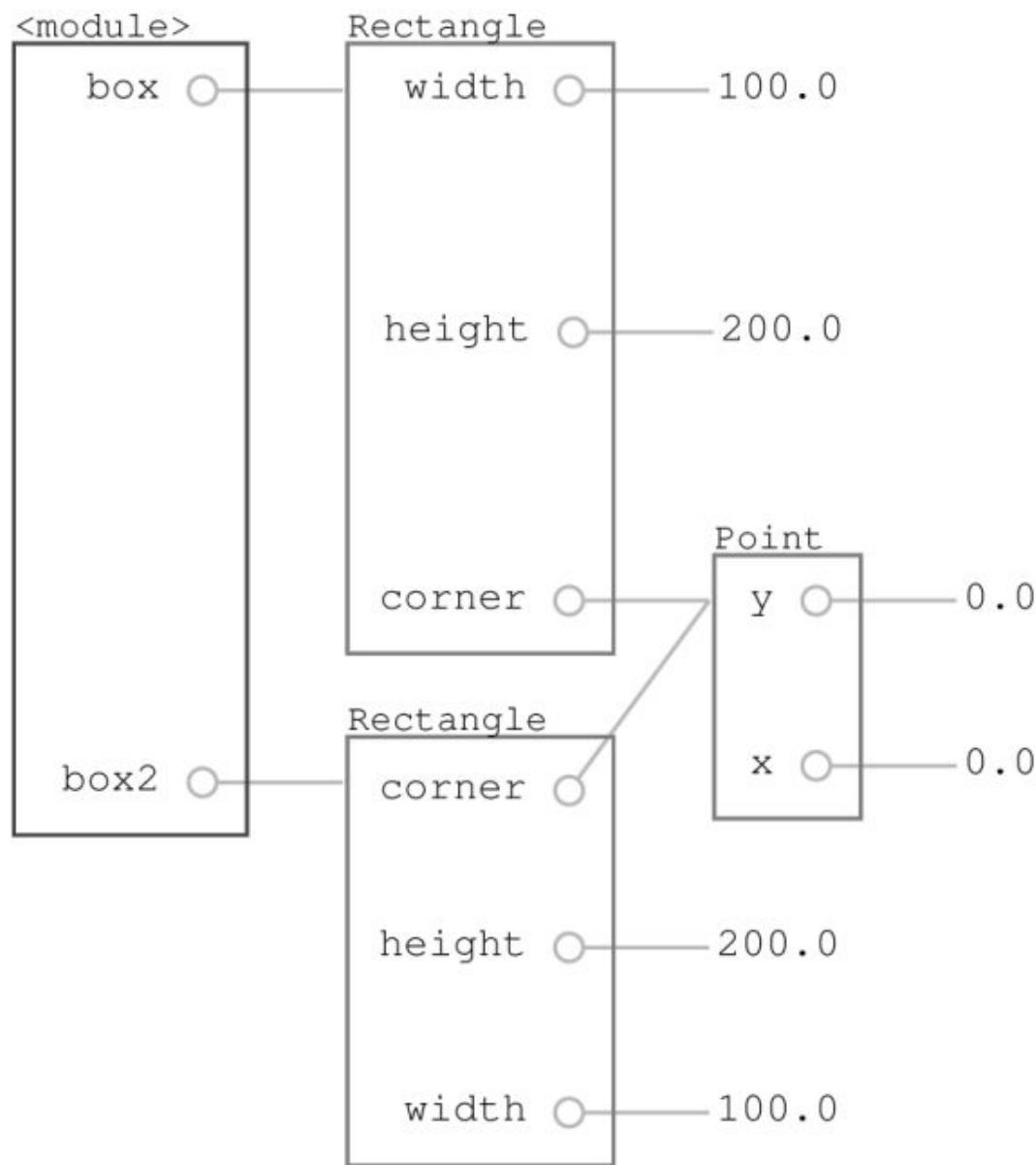
lumpy = Lumpy()
lumpy.make_reference()

box = Rectangle()
box.width = 100.0
box.height = 200.0
box.corner = Point()
box.corner.x = 0.0
box.corner.y = 0.0

box2 = copy.copy(box)

lumpy.object_diagram()
```

图C-5展示对象图的结果。`copy.copy`产生一个浅层复制，所以`box`和`box2`都有自己的`width`和`height`，但它们共享一个内嵌的`Point`对象。这种共享对于不可变对象来说通常很好，但对于可变类型，非常容易产生错误。



图C-5 对象图

## C. 4 函数和类对象

当我使用Lumpy来创建对象图时，通常会在构造引用点之前定义函数和类。那样，函数和类的定义不会出现在对象图中。

但如果传入函数和类作为实参的话，你可能希望它们在图中展示出来。下面的示例展示了如何做到，你可以从[http://thinkpython.com/code/lumpy\\_demo6.py](http://thinkpython.com/code/lumpy_demo6.py)下载它。

```
import copy
from swampy.Lumpy import Lumpy

lumpy = Lumpy()
lumpy.make_reference()

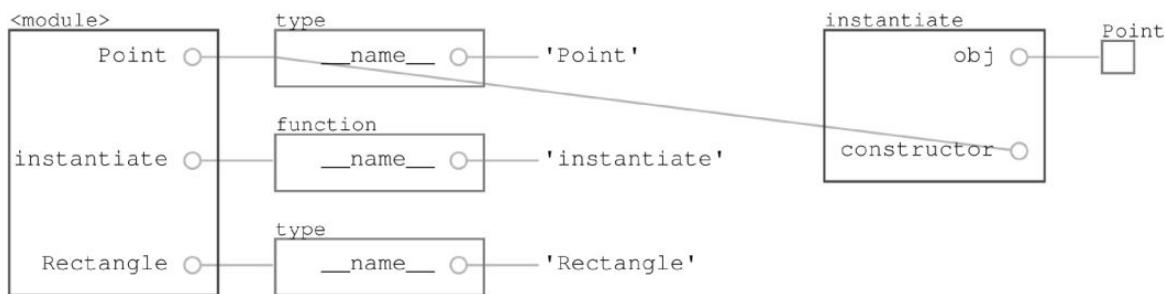
class Point(object):
    """Represents a point in 2-D space."""

class Rectangle(object):
    """Represents a rectangle."""

def instantiate(constructor):
    """Instantiates a new object."""
    obj = constructor()
    lumpy.object_diagram()
    return obj

point = instantiate(Point)
```

图C-6展示对象图的结果。由于我们在函数内部调用`object_diagram`，会得到一个栈图，里面有一帧代表模块层次的变量以及`instantiate`的调用。



图C-6 对象图

在模块层次，Point和Rectangle表示类对象（它们的类型是type）；instantiate表示一个函数对象。

这个图可以澄清两个常见的混淆点：（1）类对象Point和Point的实例obj的区别，（2）当定义instantiate时创建的函数对象，和在它被调用时创建的帧的区别。

## C.5 类图

虽然我区分状态图、栈图和对象图，但实际上它们基本上是一个东西：它们展示程序在某一时间点上的状态。

类图则不同。它们展示构成程序的类和它们之间的关系。类图是无关时间的，因为它们从整体上描述程序，而不是某个特定的时间点。例如，如果类A的一个实例通常包含了一个到类B的实例的引用，我们说这两个类之间有“HAS-A”关联。

下面是一个展示HAS-A关联的示例。你可以从  
[http://thinkpython.com/code/lumpy\\_demo7.py](http://thinkpython.com/code/lumpy_demo7.py)下载它。

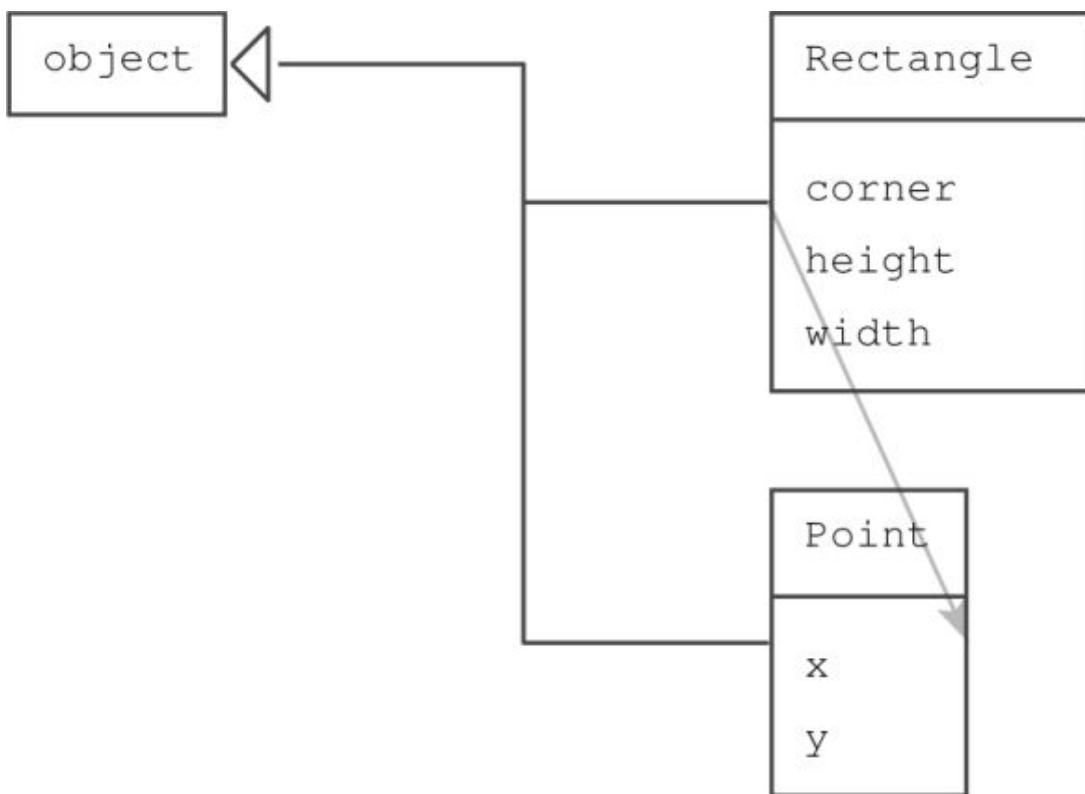
```
from swampy.Lumpy import Lumpy

lumpy = Lumpy()
lumpy.make_reference()

box = Rectangle()
box.width = 100.0
box.height = 200.0
box.corner = Point()
box.corner.x = 0.0
box.corner.y = 0.0

lumpy.class_diagram()
```

图C-7展示类图的结果。每个类用一个方盒表示，其内包含了类的名称、类提供的所有方法、所有类变量，以及任何实例变量。在这个例子中，`Rectangle`和`Point`都有实例变量，但没有方法或类变量。



图C-7 类图

从Rectangle到Point的箭头展示了Rectangle类中包含一个内嵌的Point实例。另外，Rectangle和Point都继承自object类，在图中使用一个带三角头的箭头表示。

下面是一个更复杂的示例，使用的是我对练习18-6的解答。你可以从[http://thinkpython.com/code/lumpy\\_demo8.py](http://thinkpython.com/code/lumpy_demo8.py)下载代码，你也需要<http://thinkpython.com/code/PokerHand.py>。

```

from swampy.Lumpy import Lumpy

from PokerHand import *

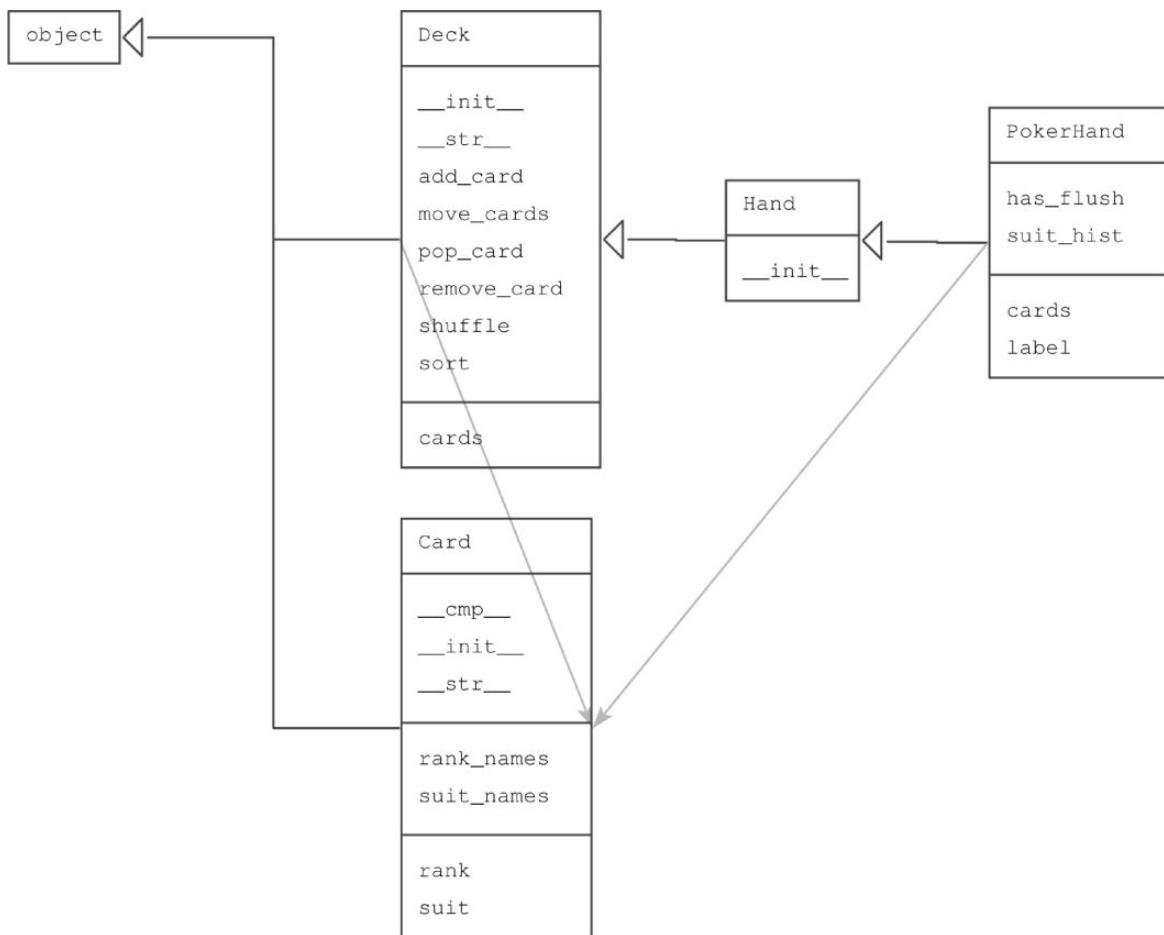
lumpy = Lumpy()
lumpy.make_reference()

deck = Deck()
hand = PokerHand()
deck.move_cards(hand, 7)

lumpy.class_diagram()

```

图C-8展示类图的结果。PokerHand从Hand继承而来，Hand继承自Deck。Deck和PokerHand都有包含Card。



图C-8 类图

这个图并没有显示Hand也包含Card，因为在程序中没有Hand的实例。这个例子显示出Lumpy的一个局限：它只知道已经实例化了的对象的属性和之间的HAS-A关联。

# 欢迎来到异步社区！

## 异步社区的来历

异步社区([www.epubit.com.cn](http://www.epubit.com.cn))是人民邮电出版社旗下IT专业图书旗舰社区，于2015年8月上线运营。

异步社区依托于人民邮电出版社20余年的IT专业优质出版资源和编辑策划团队，打造传统出版与电子出版和自出版结合、纸质书与电子书结合、传统印刷与POD按需印刷结合的出版平台，提供最新技术资讯，为作者和读者打造交流互动的平台。



# 社区里都有什么？

## 购买图书

我们出版的图书涵盖主流IT技术，在编程语言、Web技术、数据科学等领域有众多经典畅销图书。社区现已上线图书1000余种，电子书400多种，部分新书实现纸书、电子书同步出版。我们还会定期发布新书书讯。

## 下载资源

社区内提供随书附赠的资源，如书中的案例或程序源代码。

另外，社区还提供了大量的免费电子书，只要注册成为社区用户就可以免费下载。

## 与作译者互动

很多图书的作译者已经入驻社区，您可以关注他们，咨询技术问题；可以阅读不断更新的技术文章，听作译者和编辑畅聊好书背后有趣的故事；还可以参与社区的作者访谈栏目，向您关注的作者提出采访题目。

# 灵活优惠的购书

您可以方便地下单购买纸质图书或电子图书，纸质图书直接从人民邮电出版社书库发货，电子书提供多种阅读格式。

对于重磅新书，社区提供预售和新书首发服务，用户可以第一时间买到心仪的新书。

用户帐户中的积分可以用于购书优惠。100积分=1元，购买图书时，在

请输入优惠码

使用优惠码

里填入可使用的积分数值，即可扣减相应金额。

## 特别优惠

购买本电子书的读者专享异步社区优惠券。使用方法：注册成为社区用户，在下单购书时输入“57AWG”，然后点击“使用优惠码”，即可享受电子书8折优惠（本优惠券只可使用一次）。

## 纸电图书组合购买

社区独家提供纸质图书和电子书组合购买方式，价格优惠，一次购买，多种阅读选择。



**Redis实战**  
[美]约西亚 L.卡尔森 (Josiah L. Carlson) (作者)   黄健宏 (译者)   杨海玲 (责任编辑)  
分享 10 推荐 5.7K 想读 阅读

本书深入浅出地介绍了Redis的5种数据类型，并通过多个实用示例展示了Redis的用法。除此之外，书中还讲述了Redis的优化方法以及扩展方法，是一本对于学习和使用Redis来说不可多得的参考书籍。

本书一共由三个部分组成。第一部分对Redis进行了介绍，说明了Redis的基本使用方法、它拥有的5种数据结构以及操作这5种数据结构的命令，并讲解了如何使用Redis去构建文章展示网站、cookie、购物车、网页缓存、数据库行缓存等一系列程序。第二部分对Redis命令进行了更详细的介绍，并展示了如何使用Redis去构建更为复杂的辅助工具和应用程序，并在最后展示了如何使用Redis去构建一个简单的社交网站。第三部分对Redis用户经常会遇到的一些问题进行了介绍，讲解了降低Redis内存占用的方法、扩展Redis性能的方法以及使用Lua语言进行脚本编程的方法。

显示全部

● 纸质版 ￥69.00 ￥53.82 (7.8 折)  
● 电子版 ￥39.00  
● 电子版 + 纸质版 ￥55.00

现在购买 下载PDF样章

# 社区里还可以做什么？

## 提交勘误

您可以在图书页面下方提交勘误，每条勘误被确认后可以获得100积分。热心勘误的读者还有机会参与书稿的审校和翻译工作。

## 写作

社区提供基于Markdown的写作环境，喜欢写作的您可以在此一试身手，在社区里分享您的技术心得和读书体会，更可以体验自出版的乐趣，轻松实现出版的梦想。

如果成为社区认证作译者，还可以享受异步社区提供的作者专享特色服务。

## 会议活动早知道

您可以掌握IT圈的技术会议资讯，更有机会免费获赠大会门票。

# 加入异步

扫描任意二维码都能找到我们：



异步社区



微信订阅号



微信服务号



官方微博



QQ群：368449889

社区网址：[www.epubit.com.cn](http://www.epubit.com.cn)

官方微信：异步社区

官方微博：@人邮异步社区，@人民邮电出版社-信息技术分社

投稿&咨询：[contact@epubit.com.cn](mailto:contact@epubit.com.cn)