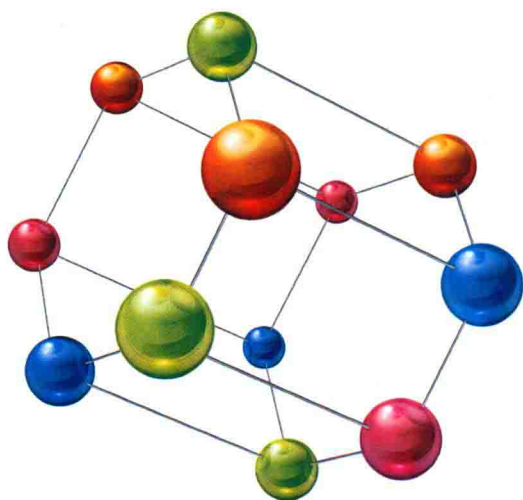


基于Python 3.7，快速掌握大数据编程工具

- 针对想直接切入Python 3及网络爬虫编程的读者
- 学习过程中贯穿大小示例，方便读者对知识点做编程实践
- 爬虫与Scrapy爬虫框架实战，提高你的综合编程能力



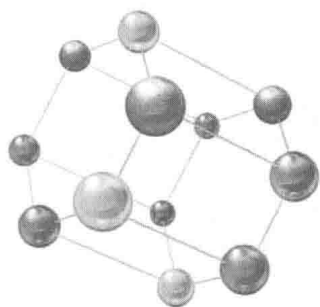
Python 3.7

编程快速入门

潘中强 薛焱 著



清华大学出版社



Python 3.7

编程快速入门

潘中强 薛焱 著

清华大学出版社
北京

内 容 简 介

随着人工智能时代的到来, Python 已经成为主流开发语言。本书就是针对完全零基础入门的读者, 采用最小化安装+极简代码的教学方式, 让读者学练结合, 达到入门 Python 与网络爬虫编程的目的。

本书分为 16 章, 从 Python 版本的演化、环境的搭建开始, 逐步介绍 Python 3.7 语言的语法基础, 包括内置类型、流程控制、函数、类和对象、异常处理、模块和包、迭代器、装饰器、多线程、文件和目录、正则表达式、网络编程、urllib 爬虫、Beautiful Soup 爬虫实战与 Scrapy 爬虫实战等。

本书概念清晰, 内容简练, 是广大 Python 与网络爬虫入门读者的佳选, 同时也非常适合高等院校和培训学校相关专业的师生教学参考。

本书封面贴有清华大学出版社防伪标签, 无标签者不得销售

版权所有, 侵权必究。侵权举报电话: 010-62782989 13701121933

图书在版编目 (CIP) 数据

Python 3.7 编程快速入门 / 潘中强, 薛焱著. —北京: 清华大学出版社, 2019
ISBN 978-7-302-51799-3

I. ①P… II. ①潘… ②薛… III. ①软件工具—程序设计 IV. ①TP311.561

中国版本图书馆 CIP 数据核字 (2018) 第 269519 号

责任编辑: 夏毓彦

封面设计: 王 翔

责任校对: 闫秀华

责任印制: 刘海龙

出版发行: 清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址: 北京清华大学学研大厦 A 座 邮 编: 100084

社 总 机: 010-62770175 邮 购: 010-62786544

投稿与读者服务: 010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈: 010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者: 三河市铭诚印务有限公司

经 销: 全国新华书店

开 本: 190mm×260mm 印 张: 18.5

字 数: 474 千

版 次: 2019 年 2 月第 1 版

印 次: 2019 年 2 月第 1 次印刷

定 价: 59.00 元



产品编号: 081508-01

前言

- Python 如何用来获取网上的数据?
- 如何分辨 Python 2.X 和 Python 3.X?
- 如何选择适合自己的 Python 版本?
- 学习 Python 用什么工具?
- 用 Windows 系统还是 Linux 系统?
- 人工智能这么火, 零基础能学 Python 吗?
- 如何用一本书学会 Python 与网络爬虫?

随着 Python 语言的普及, 越来越多非计算机专业的人们开始学习它, 而面对 Python 越来越复杂的功能, 小白读者比较迷茫, 如何学? 怎么学? 本书由浅入深, 由理论到实践, 尤其适合初级读者逐步学习和完善自己的 Python 知识结构, 最终具备自学 Python 编码的能力。本书也适合需要快速切入 Python 编程语言的技术人员。

本书特色

1. 完全零基础入门, 不需要任何前置知识

针对入门读者, 将概念通俗化地解释出来, 针对 Python 语法, 采用小示例代码演示的讲解方式, 让读者演练结合, 没有长篇大论, 无须计算机系统基础, 完全零基础入门。

2. 代码格式统一, 讲解规范

书中尽可能为每个语法都提供代码演示, 复杂内容提供详细流程。这样使得读者可以很清晰地知道每个技术的具体实现步骤, 从而提高学习的效率。

3. 循序渐进, 由浅入深

从 Python 安装到编辑器的使用, 到第一个 Python 程序, 读者每个概念每一步都可以明明白白, 中间没有任何门槛, 技术都是平滑过渡, 也非常适合自学 Python。

本书内容

第 1 章介绍 Python 的历史, 了解 Python 2.X 和 Python 3.X 的区别, 了解 Python 3.7 的变化, 然后搭建 Python 开发环境, 选择 Python 代码编辑器, 并最终实现第一个 Python 程序。

第2章简要介绍 Python 语言的一些基础知识，让读者对学习一门语言有一个概要的了解，为后面学习具体的语法铺路。

第3章介绍 Python 语言的内置类型，包括简单类型、常量类型、序列、列表、元组、字符串、字典、集合等，这些是一门开发语言的基础，正是它们构成了程序代码的最小单元。

第4章介绍流程控制和函数。它们可以帮助我们更好地管理代码，比如有些重复代码就可以放在一个函数中，这样每次只需调用函数，无须重复编码。

第5章介绍类和对象。Python 中一切皆为对象，所以了解本章就能更透彻地了解 Python 语言的基础。看完本章，读者就能看懂一点 Python 的源码了。

第6章介绍在 Python 中如何处理异常。如果要用自己的代码更安全更健壮，就必须学会异常的处理，这样当程序出错时可以更好地引导程序完成，而不是中断。

第7章介绍模块和包。很多人可能已经知道 Python 的包和模块多如牛毛，那么该如何导入别人的包、如何创建自己的包呢？学会本章，能让我们看到更多 Python 应用的可能性。

第8章介绍元类和新型类。本章会提及很多 Python 2.X 和 3.X 的区别，让读者了解 Python 中类的进化，这样就能进一步熟悉 Python 源码了。

第9章介绍 Python 迭代器、生成器、装饰器的内容。这些内容有一定的难度，但非常有用，方便代码的封装，能让代码看起来更简洁有力。

第10章介绍多线程。多线程的场景在现实中非常常见，比如双11时那么多人同时在线抢购一件商品，此时该如何处理程序呢？多线程的作用就体现出来了。

第11章介绍文件和目录。虽然我们平时的计算机操作中经常和文件、目录打交道，但是如何移动一个文件、如何添加文件的内容都需要靠代码和函数来实现。

第12章介绍正则表达式。针对零基础读者，本章详细介绍正则应用的概念、语法和原理，并演示 Python 中正则模块的各种用法。

第13章介绍网络编程。我们都经常上网，经常聊天，这些都是网络编程的功劳。本章不仅介绍网络编程的一些基础概念，还使用 Python 实现一个简单的聊天案例。

第14章介绍 urllib 爬虫。爬虫的工具很多，本章讲解的并不复杂，使用 Python 自带的 urllib 模块，演示常见的爬虫方法，其他爬虫工具其实也是基于 urllib 的，学会了它，就可以举一反三。

第15章是 Beautiful Soup 爬虫实战。读者在了解多个爬虫框架的同时，能发现 Beautiful Soup 让复杂项目变得可行，新手入门更喜欢多个框架并行研究，找到适合自己的框架。

第16章是 Scrapy 爬虫实战。前面已经学习了很多 urllib 爬虫基础，本章则让读者了解如何利用 Scrapy 框架简化自己的爬取项目工作。

代码下载

本书示例源代码下载地址可以通过扫描右边的二维码获得。

如果下载有问题，或者对本书有什么疑问，请联系电子邮箱 booksaga@163.com，邮件主题为“Python 3.7 编程快速入门”。



本书读者

- Python 与网络爬虫初学者
- Python 网络爬虫开发人员
- 其他语言转行 Python 的程序员
- 高等院校和培训学校的师生

本书第 1~12 章由平顶山学院的潘中强著、第 13~16 章由薛焱著。

著 者

2018 年 10 月

目 录

第 1 章 Python 简介	1
1.1 Python 的历史	1
1.2 为什么使用 Python	2
1.3 搭建 Python 开发环境	2
1.3.1 安装 Python	2
1.3.2 运行 Python	6
1.3.3 选择 Python IDE	8
1.4 Python 语言特性	10
1.4.1 Python 的缩进	10
1.4.2 Python 的序列	12
1.4.3 对各种编程模式的支持	12
1.4.4 Python 的动态性	12
1.4.5 匿名函数、嵌套函数	14
1.4.6 Python 自省	14
1.5 Python 2.X、Python 3.X 与 Python 3.7	15
1.5.1 Python 2.X 和 Python 3.X 的区别	15
1.5.2 Python 3.7 的新增功能	16
1.6 开始编程：第一个 Python Hello World	17
1.7 本章小结	17
第 2 章 Python 基础知识	18
2.1 Python 的基础简介	18

2.1.1	启动 Python 解释器	18
2.1.2	数值类型	19
2.1.3	字符串	20
2.1.4	列表	21
2.1.5	流程控制	23
2.1.6	函数	24
2.2	开始编程：九九乘法表	25
2.2.1	九九乘法表	25
2.2.2	编程思路	25
2.2.3	编程实现	26
2.3	本章小结	27
第 3 章	Python 的内置类型	28
3.1	Python 的类型分类	28
3.2	简单类型	29
3.2.1	布尔类型	29
3.2.2	整数类型	33
3.2.3	浮点数类型	33
3.2.4	复数类型	33
3.2.5	None 类型	34
3.3	简单类型的运算	34
3.4	常量类型	35
3.5	序列类型	36
3.6	列表类型	36
3.6.1	创建 list	36
3.6.2	list 的元素访问	37
3.6.3	列表运算	37

3.6.4	列表的方法.....	40
3.6.5	列表的内置函数 (range、filter、map)	42
3.6.6	列表推导式.....	44
3.7	元组类型.....	44
3.8	字符串类型.....	46
3.9	字典类型.....	49
3.9.1	字典的创建.....	49
3.9.2	字典的操作.....	50
3.10	集合类型.....	51
3.10.1	集合的创建.....	51
3.10.2	集合的方法和运算.....	52
3.11	开始编程：文本统计和比较.....	52
3.11.1	需求说明.....	53
3.11.2	需求分析.....	53
3.11.3	整体思路.....	53
3.11.4	具体实现.....	53
3.11.5	文本读写.....	55
3.11.6	命令行参数.....	56
3.11.7	程序入口.....	57
3.11.8	运行效果.....	60
3.12	本章小结.....	60
第 4 章	流程控制和函数.....	61
4.1	流程控制.....	61
4.1.1	选择结构.....	61
4.1.2	for 循环结构.....	63
4.1.3	while 循环结构.....	64

4.2	函数.....	68
4.2.1	函数的定义.....	68
4.2.2	函数的参数.....	68
4.2.3	函数调用和返回.....	71
4.2.4	lambda 函数.....	72
4.2.5	嵌套函数.....	73
4.2.6	函数的作用域.....	74
4.3	开始编程：八皇后算法.....	75
4.3.1	八皇后问题.....	75
4.3.2	问题分析.....	75
4.3.3	程序设计.....	76
4.3.4	问题深入.....	79
4.3.5	问题总结.....	80
4.4	本章小结.....	82
第 5 章	类和对象.....	83
5.1	面向对象.....	83
5.1.1	面向对象的历史.....	83
5.1.2	面向对象概述.....	83
5.1.3	面向对象小结.....	86
5.2	Python 类和对象.....	87
5.2.1	类的定义.....	87
5.2.2	类的实例化.....	88
5.2.3	类的方法.....	89
5.2.4	类的特殊方法.....	90
5.2.5	类的继承.....	92
5.2.6	类的关联和依赖.....	96

5.2.7	类的聚合和组合.....	97
5.2.8	类的关系.....	98
5.3	开始编程：自动打印字符图案.....	99
5.3.1	需求分析和设计.....	99
5.3.2	程序开发.....	103
5.3.3	程序入口.....	108
5.4	本章小结.....	109
第 6 章	异常捕获和抛出.....	111
6.1	异常处理.....	111
6.1.1	Traceback 异常信息.....	112
6.1.2	捕获异常.....	113
6.1.3	多重异常处理.....	114
6.1.4	异常的参数.....	117
6.1.5	内置异常类型.....	118
6.1.6	抛出异常.....	120
6.1.7	自定义异常类型.....	122
6.2	开始编程：计算机猜数.....	123
6.2.1	计算机猜数程序.....	123
6.2.2	需求分析.....	124
6.2.3	算法分析.....	125
6.2.4	编程实现.....	127
6.2.5	异常处理.....	129
6.2.6	异常类定义.....	130
6.2.7	抛出和捕获异常.....	131
6.3	小结.....	132

第 7 章 模块和包	134
7.1 模块.....	134
7.1.1 Python 模块	134
7.1.2 导入模块.....	135
7.1.3 查找模块.....	137
7.1.4 模块编译.....	137
7.2 包.....	138
7.2.1 Python 包	138
7.2.2 包的导入.....	139
7.2.3 内嵌包.....	140
7.3 本章小结.....	141
第 8 章 元类和新型类.....	142
8.1 元类.....	142
8.1.1 类工厂	142
8.1.2 初识元类.....	143
8.1.3 设置类的元类属性.....	145
8.1.4 元类的魔力.....	145
8.1.5 面向方面和元类.....	146
8.1.6 元类的小结.....	148
8.2 新型类.....	148
8.2.1 新型类和传统类的区别.....	148
8.2.2 类方法和静态方法.....	149
8.2.3 新型类的特定方法.....	150
8.2.4 新型类的特定属性.....	152
8.2.5 类的 super()方法	153
8.2.6 新型类的小结.....	154

8.3 本章小结.....	155
第 9 章 迭代器、生成器和修饰器.....	156
9.1 迭代器和生成器.....	156
9.1.1 迭代器的概念.....	156
9.1.2 生成器的概念.....	158
9.1.3 生成器 yield 语法.....	159
9.1.4 生成器的用途.....	162
9.2 修饰器.....	164
9.2.1 修饰器模式.....	164
9.2.2 Python 修饰器.....	165
9.2.3 修饰器函数的定义.....	166
9.2.4 修饰器的应用.....	167
9.3 本章小结.....	169
第 10 章 多线程.....	170
10.1 线程的概念.....	170
10.2 创建多线程.....	171
10.2.1 通过 threading.Thread()创建.....	171
10.2.2 通过继承 threading.Thread 类创建.....	172
10.3 主线程.....	173
10.4 阻塞线程.....	175
10.5 判断线程是否是活动的.....	176
10.6 线程同步.....	178
10.6.1 同步的概念.....	178
10.6.2 Python 中的锁.....	178
10.6.3 Python 中的条件锁.....	179
10.7 本章小结.....	182

第 11 章 文件与目录.....	183
11.1 文件的处理.....	183
11.1.1 获取系统类型.....	183
11.1.2 获取系统环境.....	184
11.1.3 执行系统命令.....	185
11.1.4 操作目录及文件.....	186
11.2 文件和目录的高级处理.....	192
11.2.1 复制文件.....	192
11.2.2 移动文件.....	194
11.2.3 读取压缩及归档压缩文件.....	195
11.2.4 解压文件.....	195
11.3 开始编程：文件处理实战.....	196
11.4 本章小结.....	198
第 12 章 正则表达式.....	199
12.1 正则表达式简介.....	199
12.1.1 正则表达式概念.....	199
12.1.2 正则表达式构成.....	200
12.2 re 模块的简单应用.....	201
12.3 常用正则表达式.....	205
12.3.1 常用数字表达式的校验.....	206
12.3.2 常用字符表达式的校验.....	207
12.3.3 特殊需求表达式的校验.....	209
12.4 本章小结.....	211
第 13 章 网络编程.....	212
13.1 网络编程理论基础.....	212
13.1.1 网络协议.....	212

13.1.2	IP 地址与端口	214
13.1.3	socket 套接字	215
13.2	使用 TCP 的服务器与客户端	221
13.2.1	TCP 工作原理	221
13.2.2	TCP 服务器的实现	222
13.2.3	TCP 客户端的实现	224
13.3	使用 UDP 的服务器与客户端	226
13.3.1	UDP 工作原理	226
13.3.2	UDP 服务器的实现	226
13.3.3	UDP 客户端的实现	227
13.4	开始编程：网络聊天程序	228
13.5	本章小结	231
第 14 章	urllib 爬虫	232
14.1	urllib、urllib2、urllib3 的不同	232
14.2	urllib3 中的 request 模块	234
14.2.1	对 URL 的访问	235
14.2.2	Request 类	239
14.2.3	其他类	243
14.3	request 引发的异常	243
14.4	解析 URL 的 parse 模块	244
14.4.1	URL 解析	244
14.4.2	URL 转义	248
14.5	分析 robots.txt 文件	250
14.6	本章小结	251
第 15 章	网页爬虫实战	252
15.1	获取页面源码	252

15.1.1	从网页获取数据	252
15.1.2	转换编码 UTF-8	253
15.1.3	添加关键字进行搜索	253
15.2	过滤数据	255
15.2.1	Beautiful Soup 简介	255
15.2.2	Beautiful Soup 的使用	255
15.3	数据保存	256
15.3.1	保存数据到本地文本	256
15.3.2	保存数据到数据库	257
15.4	本章小结	259
第 16 章	Scrapy 爬虫	260
16.1	安装 Scrapy	260
16.1.1	在 Windows 下安装 Scrapy	260
16.1.2	在 Linux 下安装 Scrapy	261
16.2	Scrapy 选择器 XPath 和 CSS	262
16.2.1	XPath 选择器	263
16.2.2	CSS 选择器	266
16.2.3	其他选择器	267
16.3	Scrapy 爬虫实战：今日影视	267
16.3.1	创建 Scrapy 项目	268
16.3.2	Scrapy 文件介绍	269
16.3.3	选择爬取的项目	272
16.3.4	定义如何爬取	273
16.3.5	保存爬取的结果	277
16.3.6	分派任务	278
16.4	本章小结	280

第 1 章

◀ Python 简介 ▶

Python 是一种面向对象的解释性的计算机程序设计语言，也是一种功能强大而完善的通用型语言，已经具有十多年的发展历史，成熟且稳定。Python 的语法简捷而清晰，同时有着丰富和强大的类库，可以满足日常开发方方面面的需求。

本章的主要内容是：

- 从整体上介绍 Python 语言。
- Python 语言开发环境的安装。
- Python 3.X 的特性。

1.1 Python 的历史

Python 的创始人是 Guido van Rossum。1989 年圣诞节期间，Guido 为了打发圣诞节的无趣，决心开发一个新的脚本解释程序，作为 ABC 语言的一种继承。之所以选中 Python（大蟒蛇的意思）作为程序的名字，是因为他是一个 Monty Python 的飞行马戏团的爱好者。

ABC 是由 Guido 参加设计的一种教学语言。就 Guido 本人看来，ABC 这种语言非常优美和强大，是专门为非专业程序员设计的。但是 ABC 语言并没有成功，究其原因，Guido 认为是非开放造成的。Guido 决心在 Python 中避免这一错误（的确如此，Python 与其他的语言如 C、C++ 和 Java 结合得非常好）。同时，他还想实现在 ABC 中闪现过但未曾实现的东西。

就这样，Python 在 Guido 手中诞生了。实际上，第一个实现是在 Mac 机上。可以说，Python 是从 ABC 发展起来的，主要受到了 Modula-3（一种相当优美且强大的语言，为小型团体所设计的）的影响，并且结合了 UNIX 和 C 的习惯。

虽然 Python 在国内受关注也只是近几年的事情，但是在计算机语言里面，Python 可以算是历史悠久的语言。Python 有着近 20 年的历史，版本的推进相当稳健，事实上 Python 历史比 Java 更悠久，后者是在 1991 年被设计出来的（当时名叫 oak），而真正以 Java 这个名字闻名于世则是在 1995 年。