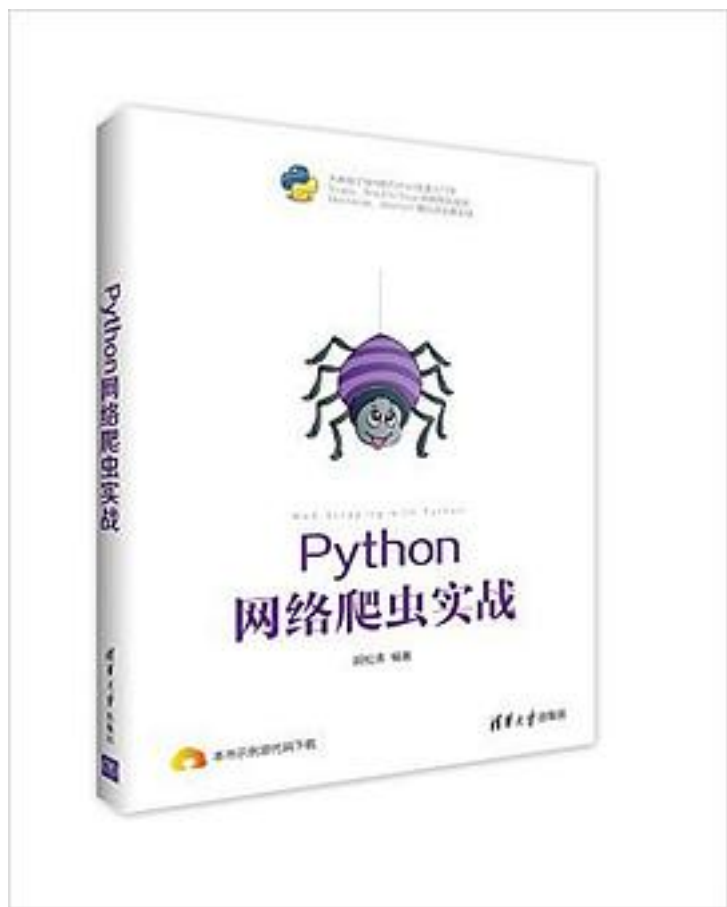


Python网络爬虫实战



[Python网络爬虫实战_下载链接1](#)

著者:胡松涛

出版者:清华大学出版社

出版时间:2016-12

装帧:平装

isbn:9787302457879

本书从Python的安装开始，详细讲解了Python从简单程序延伸到Python网络爬虫的全过程。本书从实战出发，根据不同的需求选取不同的爬虫，有针对性地讲解了几种Python网络爬虫。本书共8章，涵盖的内容有Python语言的基本语法、Python常用IDE的使用、Python第三方模块的导入使用、Python爬虫常用模块、Scrapy爬虫、Beautiful

Soup爬虫、Mechanize模拟浏览器和Selenium模拟浏览器。本书所有源代码已上传网盘供读者下载。本书内容丰富，实例典型，实用性强。适合Python网络爬虫初学者、数据分析与挖掘技术初学者，以及高校及培训学校相关专业的师生阅读。

作者介绍:

胡松涛，高级工程师，参与多个Linux开源项目，github知名代码分享者，活跃于国内著名的多个开源论坛。出版图书《树莓派开发从零开始学--超好玩的智能小硬件制作书》。

目录: 目 录

第1章 Python环境配置 1

1.1Python简介 1

1.1.1Python的历史由来 1

1.1.2Python的现状 2

1.1.3Python的应用 2

1.2Python开发环境配置 4

1.2.1Windows下安装Python 4

1.2.2Windows下安装配置pip 9

1.2.3Linux下安装Python 10

1.2.4Linux下安装配置pip 12

1.2.5永远的开始：hello world 15

1.3 本章小结 20

第2章 Python基础 21

2.1Python变量类型 21

2.1.1数字 21

2.1.2字符串 24

2.1.3列表 28

2.1.4元组 34

2.1.5字典 36

2.2Python语句 40

2.2.1条件语句——if else 40

2.2.2有限循环——for 41

2.2.3无限循环——while 43

2.2.4中断循环——continue、break 45

2.2.5异常处理——try except 47

2.2.6导入模块——import 49

2.3 函数和类 53

2.3.1函数 53

2.3.2类 59

2.4Python代码格式 65

2.4.1Python代码缩进 65

2.4.2Python命名规则 66

2.4.3Python代码注释 68

2.5Python调试 70

2.5.1Windows下IDLE调试 70

2.5.2Linux下pdb调试 73

2.6 本章小结 77

第3章 简单的Python脚本 78

3.1 九九乘法表 78

3.1.1Project分析 78

3.1.2Project实施 78

- 3.2 斐波那契数列 80
 - 3.2.1Project分析 80
 - 3.2.2Project实施 80
- 3.3 概率计算 81
 - 3.3.1Project分析 81
 - 3.3.2Project实施 82
- 3.4 读写文件 83
 - 3.4.1Project分析 83
 - 3.4.2project实施 84
- 3.5 本章小结 85
- 第4章 Python爬虫常用模块 86
 - 4.1Python标准库之urllib2模块 86
 - 4.1.1urllib2请求返回网页 86
 - 4.1.2urllib2使用代理访问网页 88
 - 4.1.3urllib2修改header 91
 - 4.2Python标准库——logging模块 95
 - 4.2.1简述logging模块 95
 - 4.2.2自定义模块myLog 99
 - 4.3 其他有用模块 102
 - 4.3.1re模块（正则表达式操作） 102
 - 4.3.2sys模块（系统参数获取） 105
 - 4.3.3time模块（获取时间信息） 106
 - 4.4 本章小结 110
- 第5章 Scrapy爬虫框架 111
 - 5.1 安装Scrapy 111
 - 5.1.1Windows下安装Scrapy环境 111
 - 5.1.2Linux下安装Scrapy 112
 - 5.1.3vim编辑器 113
 - 5.2Scrapy选择器XPath和CSS 114
 - 5.2.1XPath选择器 114
 - 5.2.2CSS选择器 117
 - 5.2.3其他选择器 118
 - 5.3Scrapy爬虫实战一：今日影视 118
 - 5.3.1创建Scrapy项目 119
 - 5.3.2Scrapy文件介绍 120
 - 5.3.3Scrapy爬虫编写 123
 - 5.4Scrapy爬虫实战二：天气预报 129
 - 5.4.1项目准备 130
 - 5.4.2创建编辑Scrapy爬虫 131
 - 5.4.3数据存储到json 138
 - 5.4.4数据存储到MySQL 140
 - 5.5Scrapy爬虫实战三：获取代理 146
 - 5.5.1项目准备 146
 - 5.5.2创建编辑Scrapy爬虫 147
 - 5.5.3多个Spider 153
 - 5.5.4处理Spider数据 157
 - 5.6Scrapy爬虫实战四：糗事百科 159
 - 5.6.1目标分析 159
 - 5.6.2创建编辑Scrapy爬虫 160
 - 5.6.3Scrapy项目中间件——添加headers 161
 - 5.6.4Scrapy项目中间件——添加proxy 165
 - 5.7scrapy爬虫实战五：爬虫攻防 167
 - 5.7.1创建一般爬虫 167
 - 5.7.2封锁间隔时间破解 171

- 5.7.3 封锁Cookies破解 171
- 5.7.4 封锁user-agent破解 171
- 5.7.5 封锁IP破解 174
- 5.8 本章小结 177
- 第6章 Beautiful Soup爬虫 178
 - 6.1 安装BeautifulSoup环境 178
 - 6.1.1 Windows下安装Beautiful Soup 178
 - 6.1.2 Linux下安装Beautiful Soup 179
 - 6.1.3 最强大的IDE——Eclipse 179
 - 6.2 BeautifulSoup解析器 188
 - 6.2.1 bs4解析器选择 188
 - 6.2.2 lxml解析器安装 189
 - 6.2.3 使用bs4过滤器 190
 - 6.3 bs4爬虫实战一：获取百度贴吧内容 196
 - 6.3.1 目标分析 196
 - 6.3.2 项目实施 197
 - 6.3.3 代码分析 205
 - 6.3.4 Eclipse调试 206
 - 6.4 bs4爬虫实战二：获取双色球中奖信息 207
 - 6.4.1 目标分析 207
 - 6.4.2 项目实施 210
 - 6.4.3 保存结果到Excel 214
 - 6.4.4 代码分析 221
 - 6.5 bs4爬虫实战三：获取起点小说信息 221
 - 6.5.1 目标分析 222
 - 6.5.2 项目实施 223
 - 6.5.3 保存结果到MySQL 226
 - 6.5.4 代码分析 230
 - 6.6 bs4爬虫实战四：获取电影信息 230
 - 6.6.1 目标分析 230
 - 6.6.2 项目实施 232
 - 6.6.3 bs4反爬虫 235
 - 6.6.4 代码分析 237
 - 6.7 bs4爬虫实战五：获取音悦台榜单 238
 - 6.7.1 目标分析 238
 - 6.7.2 项目实施 239
 - 6.7.3 代码分析 244
 - 6.8 本章小结 245
- 第7章 Mechanize模拟浏览器 246
 - 7.1 安装Mechanize模块 246
 - 7.1.1 Windows下安装Mechanize 246
 - 7.1.2 Linux下安装Mechanize 247
 - 7.2 Mechanize 测试 248
 - 7.2.1 Mechanize百度 248
 - 7.2.2 Mechanize光猫F460 251
 - 7.3 Mechanize实战一：获取Modem信息 254
 - 7.3.1 获取F460数据 254
 - 7.3.2 代码分析 257
 - 7.4 Mechanize实战二：获取音悦台公告 258
 - 7.4.1 登录原理 258
 - 7.4.2 获取Cookie的方法 259
 - 7.4.3 获取Cookie 262
 - 7.4.4 使用Cookie登录获取数据 266
 - 7.5 本章总结 270

第8章 Selenium模拟浏览器	271
8.1 安装Selenium模块	271
8.1.1Windows下安装Selenium模块	271
8.1.2Linux下安装Selenium模块	272
8.2 浏览器选择	272
8.2.1Webdriver支持列表	272
8.2.2Windows下安装PhantomJS	273
8.2.3Linux下安装PhantomJS	276
8.3Selenium&PhantomJS抓取数据	277
8.3.1获取百度搜索结果	277
8.3.2获取搜索结果	280
8.3.3获取有效数据位置	282
8.3.4从位置中获取有效数据	284
8.4Selenium&PhantomJS实战一：获取代理	285
8.4.1准备环境	285
8.4.2爬虫代码	287
8.4.3代码解释	289
8.5Selenium&PhantomJS实战二：漫画爬虫	289
8.5.1准备环境	290
8.5.2爬虫代码	291
8.5.3代码解释	294
8.6 本章总结	294
• • • • •	(收起)

[Python网络爬虫实战_下载链接1](#)

标签

Python

爬虫

python

Python爬虫

网络爬虫

骗钱

计算机

垃圾

评论

很一般，python2无感

python2.7，比较基础，大部分是我已经会的。

TP311.56 1434

大概看了一下

内容相对比较简单，Python2和Mechanize有些过时了，不少typo，总体而言尚可。

有一半是讲解python基础的，与爬虫无关。后面把流行的包或框架都讲到了，对初学者还是很不错的本书。

爬虫最重要的：bs4（静态），selenium（模拟浏览器动态），request（我是调包侠不存在的），scrapy（成熟框架） 好吧，遇到验证码这些深一点的就无能为力了……

没懂作者这个水平是怎么想出实战的书的？浪费地球资源

厉害归厉害

爬虫入门图书，喜欢这个目录结构，第一，基础知识占篇幅不大（又不是工具书，常用的列出即可），可以蹭蹭地进入实战；第二，模块分明，讲到一个装一个。缺陷是python2相对有些过时，书中打印错误也不少

入门图书

字典那一节，“字典的键可以是数字、字符串、列表、元组、几乎什么都可以”然而dict中的key只能是hashable，列表是不可以的，不推荐阅读。

一晚上翻完，小白吧，看不懂，入门以后的人呢，看着太简单。

[Python网络爬虫实战 下载链接1](#)

书评

[Python网络爬虫实战 下载链接1](#)