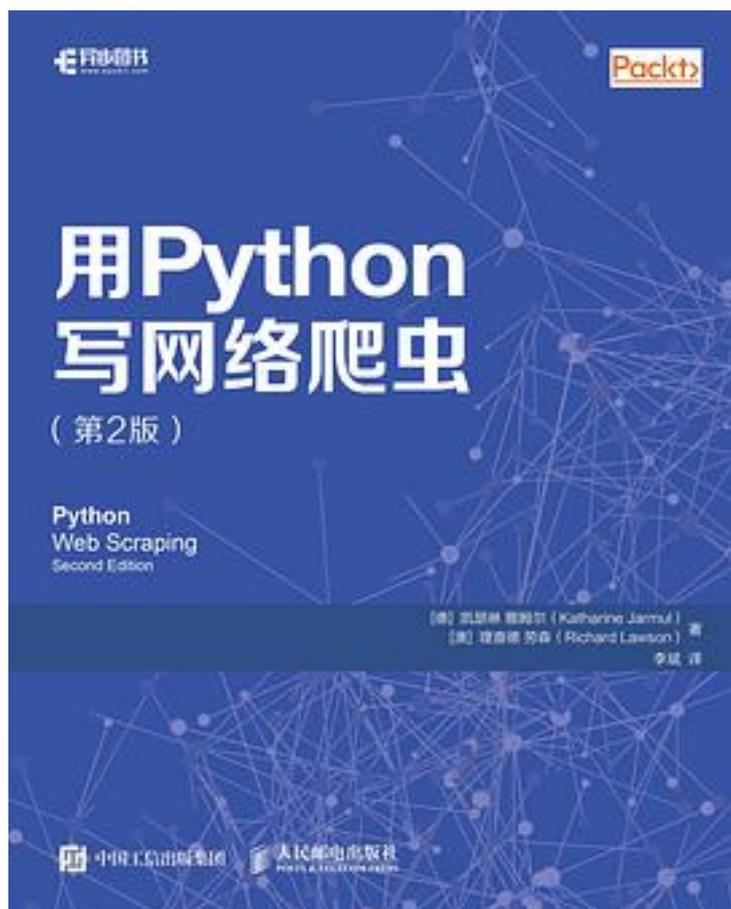


# 用 Python 写网络爬虫（第2版）



[用 Python 写网络爬虫（第2版）\\_下载链接1](#)

著者:[德] 凯瑟琳·雅姆尔 (Katharine · Jarmul)

出版者:人民邮电出版社

出版时间:2018-7-22

装帧:平装

isbn:9787115479679

本书包括网络爬虫的定义以及如何爬取网站，如何使用几种库从网页中抽取数据，如何通过缓存结果避免重复下载的问题，如何通过并行下载来加速数据抓取，如何利用不同的方式从动态网站中抽取数据，如何使用叔叔及导航等表达进行搜索和登录，如何访问被验证码图像保护的数据，如何使用 Scrapy 爬虫框架进行快速的并行抓取，以及使用

Portia 的 Web 界面构建网路爬虫。

作者介绍:

Katharine Jarmul 是德国柏林的一位数据科学家和 Python 支持者。她经营了一家数据科学咨询公司——Kjamistan，为不同规模的企业提供诸如数据抽取、采集以及建模的服务。她从 2008 年开始使用 Python 进行编程，从 2010 年开始使用 Python 抓取网站，并且在使用网络爬虫进行数据分析和机器学习的不同规模的初创企业中工作过。读者可以通过 Twitter (@kjam) 关注她的想法以及动态。

Richard Lawson

来自澳大利亚，毕业于墨尔本大学计算机专业。毕业后，他创办了一家专注于网络爬虫的公司，为超过 50 个国家的业务提供远程工作。他精通世界语，可以使用汉语和韩语对话，并且积极投身于开源软件事业。他目前正在牛津大学攻读研究生学位，并利用业余时间研发自主无人机。

译者介绍

李斌，毕业于北京科技大学计算机科学与技术专业，获得硕士学位。曾任职于阿里巴巴，当前供职于凡普金科，负责应用安全工作。热爱 Python 编程和 Web 安全，希望以更加智能和自动化的方式提升网络安全。博客地址为 [pythoner.com](http://pythoner.com)。

- 目录: 前言
- 资源与支持
- 第1章 网络爬虫简介
- 第2章 数据抓取
- 第3章 下载缓存
- 第4章 并发下载
- 第5章 动态内容
- 第6章 表单交互
- 第7章 验证码处理
- 第8章 Scrapy
- 第9章 综合应用
- • • • • ([收起](#))

[用 Python 写网络爬虫（第2版）\\_下载链接1](#)

## 标签

爬虫

python

Python

计算机

数据方法

计算机科学

計算機

Linux

## 评论

就是把整个框架了解了一边，不适合初学者，可以读一下，复习用

-----  
最近在啃这本书，发现里面的代码到处都是坑，运行起来到处报错。不过，这本书提供了爬虫的设计框架。慢慢夏日，如果能走出本书，也许会有些收获吧。

-----  
一本指导你用python写网络爬虫的工具书

-----  
内容空洞。无参考价值

-----  
更多的是偏使用，从作者的思路来看，偏CI，而不是自己hack

-----  
[用 Python 写网络爬虫（第2版）\\_下载链接1](#)

## 书评

第二段为2018年5月31日补充

本书不适合绝对入门者，适合已经熟悉python且熟悉大多数模块的人。作者对爬虫的编写考虑较为全面，且有相关练习网页可以实操。但是相关模块方法解释基本没有，整本书就是让你对爬虫认识有一个“有这么回事”的概念而不是“就那么回事”。我刚入门...

-----  
[用 Python 写网络爬虫（第2版） 下载链接1](#)