《用Python写网络爬虫》

图书基本信息

书名：《用Python写网络爬虫》

13位ISBN编号：9787115431795

出版时间：2016-8-1

作者：[澳]理查德 劳森

页数：157

译者：李斌

《用Python写网络爬虫》

内容概要

作为一种便捷地收集网上信息并从中抽取出可用信息的方式，网络爬虫技术变得越来越有用。使用Python这样的简单编程语言，你可以使用少量编程技能就可以爬取复杂的网站。
《用Python写网络爬虫》作为使用Python来爬取网络数据的杰出指南，讲解了从静态页面爬取数据的方法以及使用缓存来管理服务器负载的方法。此外，本书还介绍了如何使用AJAX URL和Firebug扩展来爬取数据，以及有关爬取技术的更多真相，比如使用浏览器渲染、管理cookie、通过提交表单从受验证码保护的复杂网站中抽取数据等。本书使用Scrapy创建了一个高级网络爬虫，并对一些真实的网站进行了爬取。
《用Python写网络爬虫》介绍了如下内容：
通过跟踪链接来爬取网站；
使用lxml从页面中抽取数据；
构建线程爬虫来并行爬取页面；
将下载的内容进行缓存，以降低带宽消耗；
解析依赖于JavaScript的网站；
与表单和会话进行交互；
解决受保护页面的验证码问题；
对AJAX调用进行逆向工程；
使用Scrapy创建高级爬虫。
本书读者对象
本书是为想要构建可靠的数据爬取解决方案的开发人员写作的，本书假定读者具有一定的Python编程经验。当然，具备其他编程语言开发经验的读者也可以阅读本书，并理解书中涉及的概念和原理。

《用Python写网络爬虫》

作者简介

Richard Lawson来自澳大利亚，毕业于墨尔本大学计算机科学专业。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。他精通于世界语，可以使用汉语和韩语对话，并且积极投身于开源软件。他目前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。

《用Python写网络爬虫》

书籍目录

目录

《用Python写网络爬虫》

精彩短评

1、刚读完，感觉讲的比较浅显，可以尽快熟悉网络爬取的流程，但深入的讲的不多，还有网上下载的源代码比较混乱，需要仔细的校正
2、介绍蛮多的爬虫技术和Python模块。其实挺不错的。最好配合培训视频来看。
3、终于读完了
4、够实用
5、适合那些没有接触过爬虫的人，想通过此书入门的话还是不错的。
全书内容不算多，很基础，比较像导论和介绍性的，更深入的内容需要自己去搜来看
6、偏简单
7、很通顺，但不够具体，这也许跟篇幅以及书本身的结构有关系。总之，看完了能明白爬虫、一整套的抓取数据流程是什么样的。有意思的是讲验证码识别部分，如果“难一点我们应该怎么样？”使用付费的人工识别服务，猝不及防！
8、很浅显的书
9、不错，挺好~
10、2.7的代码，唉:-(
11、书里面代码有bug，感觉书不太好用，没有《python 网络数据采集》条理清晰，看了两本关于爬虫的书，对比之后，高下立判。
12、0.一看标题就明白不是从入门到放弃了。
1.八天，粗率弄了一遍。得第二遍撸点儿代码了。（必须思考思考），跑跑小轮子。
13、了解网络爬虫的基本知识。
14、这本书自己敲了大概有一个月，比较简单，用一个网站介绍了一些基本的爬虫基本知识，后面的scrapy框架和例子大概看了一下算是比较不错的入门书
15、内容实用，都是比较新的工具。讲得很细，对于新手很友好。每章都提供了源码，方便学习，可以看出作者很用心。

《用Python写网络爬虫》

版权说明