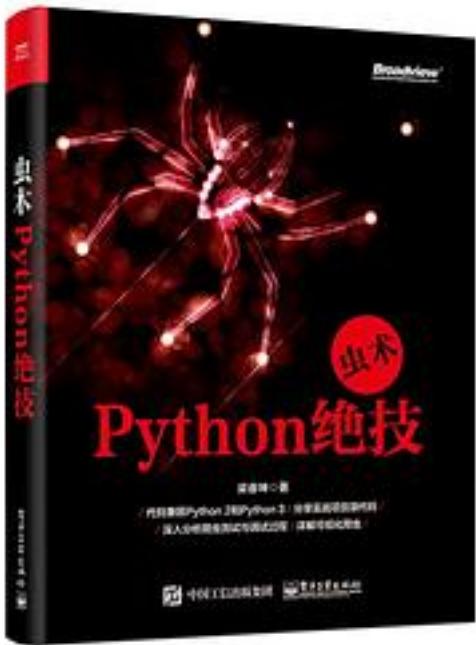


虫术——Python绝技



[虫术——Python绝技 下载链接1](#)

著者:梁睿坤

出版者:电子工业出版社

出版时间:2018-7

装帧:平装

isbn:9787121344565

《虫术——Python绝技》以大数据应用方面常用的语言 Python 为基础,从网络爬虫的实现原理入手,逐步引领读者进入网络爬虫的世界。在各类爬虫框架中,将 Scrapy 作为轴心,从多个维度揭开爬虫技术的面纱。例如,爬取规则的制定技巧,设计高速爬虫,如何让爬虫更“聪明”地获取数据,将海量数据进行分布式存储的技术,设计具有高隐匿性的爬虫,大规模、高并发的分布式爬虫技术。

《虫术——Python绝技》基于 Python 这门灵活且简洁的语言,结合作者在网络数据爬取和大数据方面的实际工程经验,使得

本书更具有实用性。《虫术——Python绝技》旨在让更多数据工作者或编程爱好者在大数据时代从海量的信息中通过掌握“虫术”来获取对自己或企业有价值的信息。

作者介绍：

近二十年软件开发、项目管理、团队建设和管理经验。致力于互联网技术应用与大数据应用方面的研究与开发工作。曾任多家软件公司的高级软件工程师、项目经理、首席架构师和技术总监等职务。

现任广州市增增智能科技有限公司CEO，从事视觉智能、语音智能及IoT等技术的产品研发与企业经营方面的工作。

目录: 第1章 爬虫初步

1.1 爬虫与大数据

1.1.1 大数据架构

1.1.2 爬虫的作用与地位

1.1.3 Python与爬虫

1.1.4 Python的网络爬虫框架

1.1.5 虫术技术路线图

1.2 实例：简单的爬虫

1.3 内容分析进阶

1.3.1 选择器

1.3.2 深入BeautifulSoup

1.3.3 元素的搜寻

1.3.4 乱码与中文编码

1.4 新闻供稿的爬取实例

1.5 小结

第2章 Scrapy基础知识

2.1 Scrapy架构

2.2 Scrapy快速入手

2.3 数据模型Item

2.4 蜘蛛—Spiders

2.5 管道—Item Pipeline

2.6 Scrapy的运行与配置

2.7 新闻供稿爬虫的Scrapy实现

2.8 小结

第3章 Scrapy的工程管理

3.1 Scrapyd

3.2 scrapyd-client及部署

3.3 搭建爬虫服务器

第4章 中阶虫术

4.1 蜘蛛的演化

4.1.1 蜘蛛的本质—深入Spider

4.1.2 通用蜘蛛

4.1.3 蜘蛛中间件

4.2 爬虫系统的测试与调试

4.2.1 开发期调试

4.2.2 蜘蛛的测试

4.2.3 蜘蛛的运行期调试

4.2.4 调试内存溢出

4.3 处理HTTP请求

4.3.1 HTTP请求

- 4.3.2 Scrapy的Request对象
 - 4.3.3 表单处理
 - 4.3.4 下载器中间件
 - 4.4 处理HTTP响应
 - 4.4.1 HTTP响应
 - 4.4.2 Scrapy的响应对象
 - 4.4.3 深入选择器
 - 4.4.4 非结构化数据的提取
 - 4.4.5 黑夜中的眼睛
 - 4.5 处理JavaScript
 - 4.5.1 示例：电商产品爬虫
 - 4.5.2 Selenium和PhantomJS
 - 4.5.3 Scrapy与Splash
 - 4.6 数据存储与后处理
 - 4.6.1 图片的下载与存储
 - 4.6.2 示例：产品图片采集
 - 4.6.3 导出到数据文件
 - 4.6.4 导出到数据库
 - 4.6.5 示例：基于阿里云的存储后端
- 第5章 高阶虫术
- 5.1 增量式爬网
 - 5.1.1 推演路由
 - 5.1.2 时机的重要性
 - 5.1.3 去重处理
 - 5.1.4 布隆过滤器
 - 5.1.5 基于Redis的布隆过滤器
 - 5.2 突破封印
 - 5.2.1 封禁浅析
 - 5.2.2 客户端仿真
 - 5.2.3 化身万千—蜘蛛世界的易容术
 - 5.2.4 反跟踪
 - 5.2.5 绕开蜜罐
 - 5.3 虫海
 - 5.3.1 分布式爬虫架构
 - 5.3.2 认识scrapy-redis
 - 5.3.3 示例：分布式电商爬虫
 - 5.4 可视化爬虫
 - 5.4.1 示例：某点评网爬虫
 - 5.4.2 解读Portia爬虫代码
 - 5.4.3 数据项加载器—Item Loaders
 - 5.4.4 最后的工作
- · · · · (收起)

[虫术——Python绝技](#) [下载链接1](#)

标签

Python

计算机

爬虫

烂

机器学习

akb

CS

评论

让我说什么好，18年的书，还在用着python2，scrapy也是基于1.0的，哎，结构也乱

首先，我是个小白，其次，该书不适合新手，应该适合小白提升，接着，由于我是个小白，只读懂书中60%的内容，其中Scrapy工程管理，处理JavaScript (Splash)，增量式爬虫等高深的内容，希望能出几个实例或是视频，让我等小白好明白一点

[虫术——Python绝技](#) [下载链接1](#)

书评

[虫术——Python绝技](#) [下载链接1](#)