

责任编辑:尹海娜  
责任印制:承伯平  
装帧设计:燕春莉



计算机类专业“互联网+”创新型精品教材

## 计算机类专业“互联网+”创新型精品教材

- ◆ 数据库应用技术(第二版) MySQL数据库项目化教程
- ◆ C语言程序设计(第二版) Python程序设计
- ◆ Java语言程序设计(第二版) 物联网工程导论
- ◆ 计算机网络技术(第二版) 大数据技术基础
- ★ 计算机组装与维护 云计算技术基础
- ★ 计算机英语(第二版) 人工智能应用导论
- ★ 办公自动化技术(第二版) 计算机组成与维护
- 网页设计与制作(第二版) Windows Server操作系统
- Linux操作系统基础与实训教程 ASP.NET网站开发技术项目化教程
- HTML5+CSS3开发实战 **数据采集技术**
- HTML5与JavaScript程序设计
- Java Web应用开发

带★标志的为“十二五”职业教育国家规划教材  
带◆标志的为“十四五”职业教育国家规划教材

数据采集技术

# 数据采集技术

主编 罗坤 薛龙 杨健

主编  
罗坤  
薛龙  
杨健



绿色印刷产品



扫描二维码  
共享立体资源

ISBN 978-7-200-18683-3

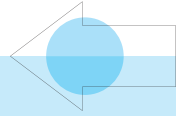


9 787200 186833 >  
定价: 42.00元

北京出版集团  
北京出版社

北京出版集团  
北京出版社

项目一 初识数据采集	1
任务一 数据采集	1
任务二 网络爬虫	11
实训 爬取手机端数据	25
项目二 采集和解析网页数据	33
任务一 采集网页分析	33
任务二 用 Python 实现 HTTP 请求	41
任务三 静态网页和动态网页采集	48
任务四 解析采集到的网页	59
实训 爬取天气预报数据	91
项目三 并行多线程网络数据采集	101
任务一 多线程网络数据采集	101
任务二 多进程网络数据采集	107
实训 爬取旅游网站数据	114
项目四 Scrapy 爬虫框架网络数据采集	125
任务一 安装 Scrapy 爬虫框架并创建爬虫项目	125
任务二 使用 Scrapy 提取网页数据	134
任务三 多层次网页抓取	142
实训 爬取图书网站数据	148



项目五 反爬虫技术与反反爬虫技术·····	161
任务一 反爬虫技术 ·····	161
任务二 反反爬虫技术 ·····	168
实训 爬取购物网站商品数据 ·····	174
参考文献·····	191

