**前言**

Python这门编程语言有着自身成熟的理论体系，同时它也是应用性极强的一门技术。

在读者从其他任何入门类型的Python图书学到Python的基本语法，读者可能会对自己所达到的这种瓶颈状态感到焦虑，不知道下一步该从哪里入手，或者说该学什么。这就是本书想要解决的问题。

在本书中，读者将会学到一些关于自动化爬虫、网页测试以及Python的拔高内容。这些东西会帮助读者构建自己的工具箱。无论读者朝向哪个方向发展，学习这些的总是不会错的。如果读者的方向是web开发，那么这些知识是必不可少的，而对于测试运维、数据分析师和人工智能工程师，那么您更需要精通这些知识。

不同于其它Python拓展性教程书，本书更注重内容拓展和迁移。将一些庞大的代码打散、拆分，转换成一种易于理解和容易掌握的内容。之所以采用这样零碎化的方式，是因为本书旨在为读者提供更为广泛的解决方案，并且读者可以在自己十万火急的项目中快速应用这些技术。笔者相信，在这书中，读者不仅可以学到很多Python的编程技能，并且可以在十万火急的任务面前快速应用它们。

本书的行文内容深入浅出、图文并茂，将枯燥生硬的知识用一种诙谐幽默、浅显直白的口语化方式娓娓道来。本书不囿于教条，更注重实战部分内容的展示。对于理论知识介绍，除了必须要了解的内容，本书绝不贪多求全，将重点放在实务操作，来帮助快速上手。

因受作者水平和成书时间所限，本书难免存有疏漏和不当之处，敬请指正。

**本书特色**

**1．内容实用实在、详略得当，讲授**

从内容结构上非常注重知识的实用性和可操作性，对于重点和难于理解的内容绝不吝惜笔墨。由浅入深的讲解方式易于读者理解，符合初学者的认知规律。更多的来说，本书采用示例与实践相结合的方式，实战性强。在书籍内容推进的同时，帮助读者建立起全套的知识理论体系

**2．三步一回头，知识综合运用**

传统的书籍注重于介绍理论知识、案例少，这样的学习学得快、忘得也快，很大程度上读者可能会出现学完后面的章节忘记了前面的章节。本书则不然，在讲解完每一个大的模块后，都会采用实际的案例对前面学习的内容进行综合处理运用，这样一来就避免了学习不同章节不同实例间的知识脱节，使读者构建的知识理论体系连贯统一。

笔者的讲授绝不是死板教条式的，尽管笔者是本书唯一的作者，但在本书中，笔者更多会使用“我们”这样的词语。原因是我们是一起的，都在朝着拓展Python技能，提高自己的技术水平的方向努力，从这一方面来说，笔者和读者不过是闻道路上互相搀扶的同道者，我走的慢了你催我一声，你陷入泥潭了我拉你一把。

**本书内容及体系结构**

第1章 从Selenium开始

本章从起初入手，帮助读者了解Selenium项目的框架和由来，以及如何选择自己的WebDriver。接着介绍了在Windows和Linux两种操作系统下的Selenium安装方式。

第2章 页面导航

本章对Selenium项目和浏览器交互的基本操作内容进行了介绍和串讲。本章介绍了如何通过Selenium项目实现启动浏览器，打开浏览器，浏览器窗口切换等一系列的常用的浏览器操作的自动化。为后面的学习奠定基础。

第3章 元素定位

在第三章中笔者将会引领读者领略了Selenium的八种元素定位方式。通过对本章与前两章的内容的学习，读者可以初步掌握并实现一个略有雏形的自动化程序的能力。

第4章 利器----PhantomJS

PhantomJS是一个独立的项目，可以实现绝大部分浏览器的功能。不同于其他的浏览器，PhantomJS只会在内存中运行，而不会显示任何界面。十分适合我们进行数据采集的时候使用。本章介绍了PhantomJS和Selenium搭配使用的方法。

第5章 实战part1——Python官网

本章是对前面所学的内容的一个综合整理运用。通过实战，帮助读者巩固已经学习过的知识，并且在实际操作中，锻炼读者的编程能力与思维。

第6章 实战part2——今日头条

以今日头条为例子，引入即将要学习的内容，本章主要讨论技术部分实现以及解决思路分析。目的在于帮助读者形成一个目录式的系统印象。

第7章 数据编解码、处理

在互联网中，难免要面对错综复杂的各种数据类型，这一章主要介绍以下内容：CSV、JSON、XML、BASE64、NLTK、HASH、Bloom Filter以及关系型数据库和非关系型数据库。

第8章 实战part3——猫眼电影

以猫眼电影为例子，详细讲解从了解客户需求到网站分析、处理的全套流程，并在分析中逐步引入代码的说明与实现。

第9章 实战part4——淘宝商品

在这一小节中，笔者以淘宝商品为例子，阐述如何应对复杂网站的页面数据采集，帮助诸位追赶较为新颖的实现方式，解决读者因难以下手而犹豫不决，无法直接应用的难题。

第10章 单元测试

解决单元测试是什么、为什么要写单元测试、应该怎么写单元测试这三个难题。帮助读者评估在自己的项目中是否要使用单元测试。

第11章 多线程

实现同步完成多项任务的功能，帮助读者提高程序的运行效率，解决时间。讨论线程、进程、线程切换、锁以及GIL等概念。

第12章 发送邮件

本章讨论程序反馈最低廉且合理的方式，通过实现自动化发送邮件，实现服务器（程序）无人值守情况下的报备以及记录

第13章 Selenium IDE

介绍了两个简单且方便的录制自动化脚本的工具，为读者提供一条功能化的测试道路，快捷生成自动化脚本。

第14章 Python拓展

这一章节讨论包括但不限于：Python2代码转换Python3代码，测试类型，通配符，字符串方法，异常层次结构，类，推导式等等扩展性内容。

第15章 GUI，图形化测试

讨论制作图形化界面的方法并且介绍经常使用的组件，帮助读者实现测试工具图形化以及了解底层实现的思路。

第16章 实战案例part5 知乎

着重讨论Selenium项目和传统方式的爬虫的结合，实现一种全新的，功能更加强大的数据采集爬虫。

第17章 实战案例part6 微博

对全书的内容的整合提升，帮助读者巩固已经学过的知识，构筑全面的知识框架，了解自己的不足与缺陷。

**本书读者对象**

* 测试自动化
* Web开发者
* Python开发工程师
* 人工智能工程师
* 爬虫开发工程师
* 数据分析工程师
* 系统运维工程师