

Python 进阶训练营

尹会生

尹会生

极客时间《零基础学 Python》《Linux 实战技能 100 讲》视频课程讲师。
曾任新浪网研发中心技术经理和某著名培训机构高级讲师。
曾参与编写《白话大数据与机器学习》《运维前线》等书籍。

工作中使用 Python 开发了大数据采集和分析平台，使用 Django 实现了公司基础架构统计工具，同时还在使用 Python 与 TensorFlow 进行游戏、大数据以及深度学习相结合的研究工作。

① 课程介绍、学习方法与简单爬虫

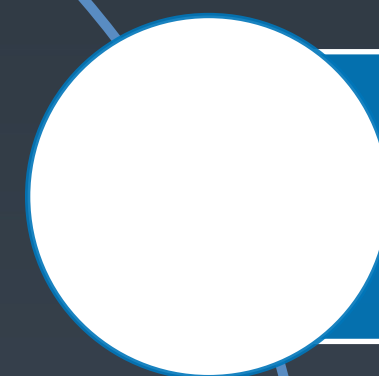


目录

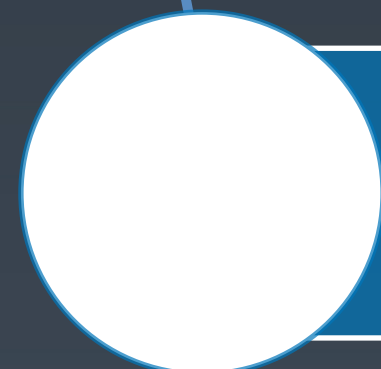
CONTENTS

- 01 课程介绍
- 02 绪论
- 03 提升 Python 开发效率的工具
- 04 复习 Python 的基础语法
- 05 使用 Python 库获取豆瓣影评和书评

学习目标



掌握高效系统学习 Python 的方法



提升 Python 开发效率的工具



复习基础 Python 语法

01 课程介绍

课程介绍

1. 完整案例：NLP 舆情系统

需求描述-数据收集-数据处理-语料处理-深度学习-结果测评-调参-展示

2. Python 底层功能

作用域问题

面向对象编程

多线程、协程

爬虫框架、Web 框架

02 绪论

为什么学习 Python

1. Python 语言足够流行
2. Python 语言足够 “简单”
3. 生态完善

Python 作为第二语言

1. 作为第二语言可以做知识迁移
2. 前端开发可以从爬虫项目开始无痛入手
3. 后端开发可以对比静态语言和动态语言的差异

高效学习 Python 的方法

1. 建立高效的学习模型
2. 了解 Python 的长处
3. 了解 Python 的特性
4. 多看高手的代码 (GitHub)
5. 好的问题是成功的一半 (Google、Stack Overflow)
6. 风格指引 (PEP8、Google Python Style Guides)

03 提升 Python 开发效率的工具

提升 Python 开发效率的工具

1. Visual Studio Code: 高效的 IDE
 - `pylint`
 - `autopep8`
 - `remote-ssh`
2. Python3.6 or Python3.7 : 不同的版本差别在哪里 ?
3. Jupyter Notebook: 数据科学家的最爱

从一个需求开始

获取豆瓣读书 Top 250 的书籍名字和评分

<https://book.douban.com/top250>

实现

步骤：

1. F12 调试模式分析网页源代码
2. 用 Requests 爬取网页全部内容
3. 用 Beautiful Soup 解析网页提取关键信息
4. 用 csv 文件存储书籍名字和评分

pip 安装加速

国内常见的镜像站：

豆瓣：<http://pypi.doubanio.com/simple/>

清华：<https://mirrors.tuna.tsinghua.edu.cn/help/pypi/>

升级 pip：

方法一：

```
pip install -i https://pypi.tuna.tsinghua.edu.cn/simple pip -U
```

方法二：

```
pip config set global.index-url http://pypi.doubanio.com/simple/  
pip install pip -U
```


pip 安装加速

配置文件：

windows : c:\Users\xxx\pip\pip.ini

Linux: ~/.pip/pip.conf

配置文件格式：

[global]

index-url = <https://pypi.tuna.tsinghua.edu.cn/simple>

翻页都是怎么做到

https://search.bilibili.com/all?keyword=%E7%8E%8B%E8%80%85%E8%8D%A3%E8%80%80%E9%B2%81%E7%8F%AD%E5%A4%A7%E5%B8%88&from_source=nav_search_new

https://search.douban.com/book/subject_search?search_text=python&cat=1001

https://weibo.com/a/aj/transform/loadingmoreunlogin?ajwvr=6&category=0&page=7&lefnav=0&cursor=&__rnd=1575299380621

https://weibo.com/a/aj/transform/loadingmoreunlogin?ajwvr=6&category=0&page=8&lefnav=0&cursor=&__rnd=1575299407336

格式化字符串

三种常用的格式化字符串方式：

1. % 操作符
2. str.format(*args, **kwargs)
3. f-string：Python 3.6 引入，该方法源于 PEP498

格式化字符串

% 操作符 -- printf 风格的字符串格式化

```
import math  
print('The value of Pi is approximately %5.3f.' % math.pi)  
# 输出：The value of Pi is approximately 3.142.
```

格式化字符串

.format -- 更加灵活

```
print('{1} and {0}'.format('spam', 'eggs'))
```

输出： eggs and spam

```
print('The story of {0}, {1}, and {other}'.format('Bill', 'Manfred', other='Georg'))
```

输出： The story of Bill, Manfred, and Georg.

参考： <https://docs.python.org/zh-cn/3.6/library/string.html#formatstrings>

格式化字符串

f-string : Python 3.6 引入 , 该方法源于 PEP 498。

<https://docs.python.org/zh-cn/3.6/whatsnew/3.6.html#whatsnew36-pep498>

f-string 和 %操作符、.format 比较 :

1. 性能更好
2. 易读性好

三种写法比较

```
firstname = 'yin'
lastname = 'wilson'
print('Hello, %s %s.' % (lastname, firstname))
print('Hello, {1} {0}.'.format(firstname, lastname))
print(f'Hello, {lastname} {firstname}.')
```

格式化字符串

f-string还可以做其他事情：

```
f'{ 2 * 5 }'
```

```
class Person:
    def __init__(self, first_name, last_name):
        self.first_name = first_name
        self.last_name = last_name

    def __str__(self):
        return f'hello, {self.first_name} {self.last_name}.'

    def __repr__(self):
        return f'hello, {self.first_name} {self.last_name}.'

me = Person('yin', 'wilson')
print(f'{me}')
```

04 复习 Python 基础知识

复习 Python 基础知识

赋值使用 “=” 符号

Python 3.8 会引入海象运算符 “:=” (PEP572)

赋值前不需要指定类型（分配内存）

数值、布尔、字符串、序列赋值前不需要指定类型（分配内存）

复习 Python 基础知识

基本数据类型	
None	空对象
Bool	布尔值
数值	整数、浮点数、复数
序列	字符串、列表、元组
集合	字典
可调用	函数

复习Python基础知识

循环控制	
条件语句	if...else
循环语句	for...in , while
导入库、包、模块	import

注意：Python 使用缩进作为语句块的分隔

复习 Python 基础知识

Python 的执行方法：

```
shell> python filename.py
```

Python 会将 .py 文件编译成字节码 pyc 格式文件，由 Python 虚拟机执行

```
shell> python ( 回车 )  
python> import whatyouwant  
python> run python something
```

交互模式（解释模式）

总结

1. 通过静态页面爬虫复习了 Python 的基础语法
2. 了解 Python 语言作为动态语言的特点
3. 掌握静态页面数据收集的一般方法

THANKS! |  极客大学