https://www.icourse163.org/topics/python-sp

### 利用爬虫技术能做到哪些很酷很有趣很有用的事情？

**利用爬虫技术能做到哪些很酷很有趣很有用的事情？**

准备学习python爬虫。各位大神都会用爬虫做哪些有趣的事情？ 今天突然想玩玩爬虫，就提了这个问题。跟着YouTube上的一个tutor写了个简单的程序…显示全部

关注者

**74,488**

被浏览

**6,957,984**

关注问题写回答

​62 条评论

​分享

​邀请回答

​

**655 个回答**

默认排序​

[Pacino](https://www.zhihu.com/people/ljx95)

[**Pacino**](https://www.zhihu.com/people/ljx95)

忙于追求更好的人生。

3,644 人赞同了该回答

我当初是看到这个帖子才知道Python这门语言的功能，才开始去学的，现在也学了一小段时间。不得不说，Python爬虫对于我来说真是个神器。之前在分析一些经济数据的时候，需要从网上抓取一些数据下来，想了很多方法，一开始是通过Excel，但是Excel只能爬下表格，局限性太大了。之后问了学编程的朋友，他说JavaScript也能实现，于是懵懵懂懂地就去学Java（我那朋友是学Java的，我当时问他用Java能不能实现，他说JavaScript好像可以，当时我什么都不懂，就把JavaScript理解成是Java下的一个分支，以为JavaScript只是Java其中一个包什么的，于是我便去学了一小会Java，无知惹的祸啊。。。）。但整个Java体系也太庞大了，学起来力不从心，毕竟我只是要运用其中一部分功能而已，于是学没多久我就放弃了。就在我迷茫的时候，我发现了Python大法……

废话说多了，说说自己的学习经历吧。也给想学Python，想写爬虫的人一个参考。  
一开始我是在网易云课堂上自己找了个基础的视频来学，Python是真是门简单的语言，之前懂一点Visual Basic，感觉Python也很适合给无编程基础的人学习。  
入门视频到最后，就做出了我的第一个爬虫——百度贴吧图片爬虫（相信很多的教程都是以百度贴吧爬虫为经典例子来说的。）

一开始代码很简单，只能爬取第一页的数据，于是我加了一个循环，就能够爬取制定页数的图片了。并且图片是有按顺序排列的，非常方便。在筛选网址的时候用正则表达式就好了。正则表达式使用：[[精华] 正则表达式30分钟入门教程](https://link.zhihu.com/?target=http%3A//www.oschina.net/question/12_9507)

### 【Python数据分析与展示】（八）数据合并、规整化

<https://blog.csdn.net/polarislove36/article/details/78847869>

【Python数据分析与展示】（八）数据合并、规整化

### Python数据分析与展示

**课程大纲**

【第〇周】数据分析之前奏

“数据分析”课程内容导学

Python语言开发工具选择

Anaconda IDE的基本使用方法

【第一周】数据分析之表示

第1周课程导学

单元1：NumPy库入门

单元2：NumPy数据存取与函数

单元3：实例1：图像的手绘效果

第一周测验：数据分析之表示

【第二周】数据分析之展示

第2周课程导学

单元4：Matplotlib库入门

单元5：Matplotlib基础绘图函数示例（5个实例）

单元6：实例2：引力波的绘制

第二周测验：数据分析之展示

【第三周】数据分析之概要

第3周课程导学

单元7：Pandas库入门

单元8：Pandas数据特征分析

单元9：实例3：房价趋势的关联因素分析

第三周测验：数据分析之概要

【第四周】数据分析之处理

第4周课程导学

单元10：数据的清洗和规约

单元11：时间序列的处理和展示

单元12：商品数据时序分析图

第四周测验：数据分析之处理

**预备知识**

本课程需要学习者具备Python语言编程的基本知识和初步技能，建议Python零基础学习者先修嵩老师的“[Python语言程序设计](https://www.icourse163.org/course/BIT-268001" \t "_blank)”课程。具体地，学习者需要预先掌握Python的数字类型、字符串类型、分支、循环、函数、列表类型、字典类型、文件和第三方库使用等概念和编程方法。

**证书要求**

本课程采取百分制，60分-79分可获得合格证书，80分以上可获得优秀证书。

**参考资料**

**Python集成开发环境(IDE)**

[1] Anaconda: <https://www.continuum.io/> （推荐）

[2] IDLE: Python解释器默认工具

[3] PyCharm: <https://www.jetbrains.com/pycharm/>

**参考教程**

[1] Python零基础入门教程:《Python语言程序设计基础(第2版)》，嵩天、礼欣、黄天羽著，高等教育出版社，2017.2

教程链接：[京东自营旗舰店](https://item.jd.com/12128326.html?dist=jd" \t "_blank)  [高教社天猫旗舰店](https://detail.tmall.com/item.htm?spm=a1z10.1-b-s.w19095786-17077471552.4.761f613c29xVTL&id=562950803430)  [亚马逊地址](https://www.amazon.cn/Python%E8%AF%AD%E8%A8%80%E7%A8%8B%E5%BA%8F%E8%AE%BE%E8%AE%A1%E5%9F%BA%E7%A1%80-%E5%B5%A9%E5%A4%A9-%E7%A4%BC%E6%AC%A3-%E9%BB%84%E5%A4%A9%E7%BE%BD/dp/B06W9KM5P5/ref=sr_1_1?s=books&ie=UTF8&qid=1488437402&sr=1-1)

[2] 专题参考资料：《利用Python进行数据分析》，Wes McKinney著，O’Reilly & 机械工业出版社，2014.1（该书使用Python 2.x系列，内容略微陈旧，**仅做参考**，**不建议跟踪学习**）

**参考网站**

[1] Python NumPy: <https://www.numpy.org/>

[2] Python Matplotlib: <https://matplotlib.org/>

[3] Python Pandas: <https://pandas.pydata.org/>

**常见问题**

**Q1：**除了Python，这个课程需要其他编程语言基础吗？

**A1：**

不需要，只要具备基本的Python编程能力，就可以学习本课程。

**Q2：**Python 2.x和Python 3.x，这个课程采用哪个版本？

**A2：**

Python 3.x已经足够成熟，这是Python语言的现在和未来，嵩老师所有Python课程都采用Python 3.x系列版本。

**Q3：**在线开放课程看不到老师，有问题谁来解答？

**A3：**

为了更好服务同学们，本课程教师和助教会经常在线答疑，尽快解决与课程相关的各类问题。

**Q4：**有没有好的参考书推荐？

**A4：**

本课程提供了视频、教学资源，但暂时没有参考书推荐。如果大家发现该领域优秀的参考书，可以反馈给老师

## \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*