

Python爬虫系列

专栏简介

本专栏旨在让刚刚完成Python入门的朋友，能尽快用学习到的Python语言来完成一些工作。而爬虫正好非常适合这一点。它可以让开发者快速获得成就感，而且我个人认为它的门槛并不是很高，当然如果你想成为这一领域的大神，通过本专栏是实现不了的啦。

Python网络爬虫工程师需要掌握的最少技能如下：

- Python（本专栏使用的是Python3）
- HTML·CSS·Javascript等前段知识。
- 网络服务相关内容（WEB/APP等常见的网络请求方式。）
- 数据库（Mysql/MongoDB）基本操作（增·删·改·查）

介于以上这几点，本专栏的基本大纲如下（实际更新可能会有所不同，只会多不会少。）

1. Python环境的搭建/Python语法快速复习。
2. HTML/JavaScript相关介绍。
3. 常用的爬虫库的简介。
4. 静态网站的爬取（图库，PPT等。）
5. 动态网站的爬取。
6. 数据库的操作（可能会在每个爬虫案例中穿插讲解。）
7. 一些常见的反爬策略的对应。
8. 与其他技能的结合（Django等，这一部分打算做长期更新。）

主要因为现在单纯的爬虫工作可能不是很好找，如果能和其他的框架结合起来，会更好一些，毕竟爬虫只是用来获取数据的。而如何使用这些数据才是最重要的。

关于本专栏的作者以及更新方式。

大致介绍了本专栏的内容后，接下来介绍一下我自己，以及学习本专栏的注意事项。

作者

Tango，非科班出身的编程爱好者，所以如果文章中有什么写的不对的地方，还希望大家多多指正和包涵。目前就职于NTT DATA（中国）信息技术有限公司（NCIT）。在公司主要用C#给部门内部做一些自动化工具等。平时有时间在B站做一些关于编程的直播（地址见下文）。

专栏的更新形式

本次专栏已公众号文章（首发）加视频直播讲解的方式进行。直播的视频会发布到B站或者网易云课堂等网站。

更新频率为每周六发布公众号文章，周日视频直播答疑。答疑视频的录播基本次日就能看到。

专栏资源

- IT蜗壳：www.it-worker.club 我的个人博客（很久没有更新...）
- 网易云课堂：<https://study.163.com/provider/400000000656005/course.htm>
- B站：<https://space.bilibili.com/397260706>
- GitHub：<https://github.com/it-worker-tango/bilibili>
- 我的微信：it-worker_club（请备注一下来自公众号或者目的）

注意事项

本专栏的目的是和大家分享一些爬虫相关的技术，专栏中涉及到网站以及技术请大家合理使用。不要给网站所有者造成困扰。

最后，希望大家能在本专栏中有所收获。也希望大家能和我多多交流，如果你觉得专栏的文章对你有帮助希望你能分享给你的朋友。谢谢大家的支持。