工作日记20171223动态页面抓取

Ringo Yungpo Kao

December 23, 2017

# 动态页面抓取

## 内容

由于AJAX技术的盛行，使得爬虫需要用另外一种方式去爬取，不过也是可以爬取的，只要找到了链接，剩下的就是解析json数据了。动态页面的数据都是完整而且统一的，所以只要获取得到，那处理那一部分基本不用操心。

## 基础知识的阅读

首先当然是要对基础知识有一个了解，对于AJAX链接的获取主要是通过浏览器的控制台来获取，而AJAX局部刷新的特点使得交互需要通过json来解析，这时就可以用json这个库来对json数据进行获取。

## 练习

对于基础知识了解完后，就要小小地练习一下。跟着书上地例子，把作者地网站评论内容爬取了一下。通过练习，实现了将评论内容显示出来的功能。既然已经了解了静态和动态的内容抓取，那对于数据的抓取应该没有多大问题，只是这过程还是要通过人工半自动地去判断哪个链接，这样未免不够自动化，希望将来地学习能够解决。