爬取网页：信用广东网行政处罚

网址：<http://www.gdcredit.gov.cn/infoTypeAction!xzTwoPublicListIframe.do?type=7>



每页有10条链接，链接内容如下：



要求：将图中表格的每条数据抓取下来存入数据库

数据库表名：t\_guangdong

每条数据的20个数据项的的名称：case\_no ,case\_name,punish\_type1,punish\_type2,punish\_reason,law\_item,punish\_result,entity\_name,credit\_no,org\_code,reg\_no,tax\_no,identity\_card,legal\_man,punish\_date,punish\_agent,area\_code,current\_status,offical\_updtime,note

在每条数据前加上编号id

验收条件：

1 完整的代码（需要适当的注释）,爬取的数据保存到到数据库，最少要求爬取200条数据到数据库中。

2 注意一定要使用给定的sqlalchemy数据库框架

3 要确保所编写程序的稳定性，调节合适的访问频率（注意网站的反爬取机制），以确保能够实现连续的爬取，程序不中断（重点）。

4 数据获取考核任务考核期为30天

语言选择：python（编译器推荐pycharm）

推荐入门学习材料：

极客学院Python初级课程：会员课程-知识体系图-Python课程

廖雪峰的博客：推荐看完函数部分即可，后面有选择的看。

百度：你要相信百度上的东西还是很多的