**中国执行信息公开网爬虫测试文档**

一、测试任务概述

1.测试目标

能够成功扒取中国执行信息公开网中被执行人姓名/名称、身份证号码（不全）/组织机构代码、执行法院、立案时间、案号、执行标等。

2.测试环境：

（1）操作系统：Windows 10

（2）测试工具：Visual Studio Code、Python 3.6 IDLE、Google Chrome

二、测试用例

1.测试用例：测试验证码图片获取情况

测试结果：获取验证码图片连接成功；

2.测试用例：下载验证码图片的情况

测试结果：验证码图片下载成功；

3.测试用例：获取信息列表；

测试结果：信息列表获取成功，并且每个列表有10个人的信息；

4.测试用例：获取查询结果的个人的所有信息

测试结果：获取成功，结果如下：



5.测试用例：验证码图片处理情况

测试结果：使用图像处理函数fall、ganraoxian、two对验证码图片进行去除噪点、去除干扰线（如果存在干扰线）、二值化（生成灰度图）的操作，结果如下：



6.测试用例：使用pytesseract模块的image\_to\_string函数识别验证码中的字符串

测试结果：识别正确率为0

7.测试用例：使用Github上的验证码开源识别框架cnn\_captcha来识别验证码

测试结果：由于机器配置问题，识别正确率大概在5% - 10%

8.测试用例：使用聚合数据验证码识别API：https://www.juhe.cn/docs/api/id/60

测试结果：由于中国执行信息公开网从6月10日开始一直都是502无法访问的状态，所以无法获取验证码进行识别。