

猿人学-爬虫对抗比赛事宜

写在最前面的话【很重要】

建了一个 本网站的 刷题 交流群，扫下方二维码，备注“刷题”，加我私人微信，拉你进 猿人学-爬虫对抗赛 刷题 交流群。后面新出的对抗题，也在群里通知。



王平 

阿鲁巴



扫一扫上面的二维码图案，加我微信

刷题规则（原比赛限制规则）

一、刷题禁用技术(非常重要，如果想通过刷题提升能力，烦请细读)

首先，我们鼓励进行一些骚操作进行解题。如 SwitchyOmega, mitmdump, sekro 等。从而为解题提供更多的思路。

但是，在最开始比赛的过程中，由于比赛性质偏向于对抗，所以对一些爬虫通杀手段进行了限制。限制的目的也很单纯，避免由于骚操作失去了比赛原有的意义。由于比赛已经结束，所以对于技术的限制自然已经解除。但是官方依然建议刷题时严格遵守下面的刷题规则，否则将无法发挥/体验到题目本身的魅力。爬虫的水平也无法得到应有的提升。下面是对于技术的使用的限制：

1. 比赛禁止使用 selenium / pyppeteer / puppeteer 等一系列驱动浏览器自行渲染的工具进行抓取（包括无头浏览器在内）严禁使用 SwitchyOmega, mitmdump 等手段依托于浏览器进行数据采集。

2. 比赛禁止使用 sekro 等 websocket 框架，严禁利用 jsRPC 进行抓取。严禁 OCR 等非爬虫操作（第八题的验证码可以使用 OCR 识别）！

3. 比赛中允许使用 execjs/pyv8/node.js 接口等方式去进行 js 代码的执行，但在执行的 js 代码中禁止导入一切第三方库（如：jsdom 等），若有自己写的 node.js 库，需将核心逻辑写入该执行的 js 文件中。

4. 比赛允许在抓取调试过程中使用 AST 辅助 js 进行分析。原则上可以搭建 node.js 接口（即利用 express）然后调用接口处理 js。

5. 由于题目并没有对 IP 进行任何限制（并没有必要进行限制），所以比赛中禁止使用 IP 代理池、验证码打码平台等第三方提供的爬虫辅助服务。

6. 由于比赛中对抓取速度没有要求，所以禁止使用 scrapy 等爬虫框架以及多线程、多进程、异步等加快抓取速度的采集方法。

二、刷题注意事项

刷题时，调试代码推荐使用 **chrome** 浏览器（并且**关闭** chrome 相关插件）并且**校准系统**的时间，其他浏览器可能会引发各种异常问题导致网页无法正常显示。题目已在多台电脑的 **chrome 无痕浏览器**测试正常。不保证对其他浏览器的兼容性。

三、写在最后的话

比赛结束后有太多的朋友希望比赛平台能够继续保持开启，作为一个长期爬虫练习靶场供大家提升自己。所以最后决定保持平台的继续开放。题目也会**不定期的**进行**更新**。做出这个决定也是能让朋友们尽可能多的见识到市面上出现的常规反爬手段，并且不必有**后顾之忧**的去练习爬虫的水平（或者是写解题技术博客给大家提供思路）、从而让您在真正的爬虫工作中如鱼得水，来去自如。提升自身市场的竞争力。为了最大程度的**提升您的爬虫**能力，我们依然**建议**您刷题时能够遵守以上的刷题规则，从而提升自身水准。

最后，如果您有任何的意见或者好的出题建议。或者对于题目的解题思路有任何疑问，欢迎您扫描本文件**最上方**的二维码并备注“刷题”，拉您进 **猿人学-爬虫对抗赛 刷题 交流群**，与大家共同进步。

生而为虫，我很快乐！