爬虫概览

鬼大师 @ 202105

基本流程

分析 10% > 突破爬 40% > 解析 40% > 存储 5% > 调度 5%
本期主要分享“爬”

以上流程针对于小量爬取需求，调度与存储简单的情况

一些基本必要概念

HTTP(S), HEADER, COOKIE, AJAX
代理, 网络监听
HTML, JSON, JS
Headless Browser
反爬

5分钟预售券抢购爬虫

预售券抢购 charles, ajax, node

就这? 误解?

开始战斗

各种技术的优缺点，仅供娱乐

裸奔

普通请求分析结果

单因子筏值封禁 如封ip
生成复合标识

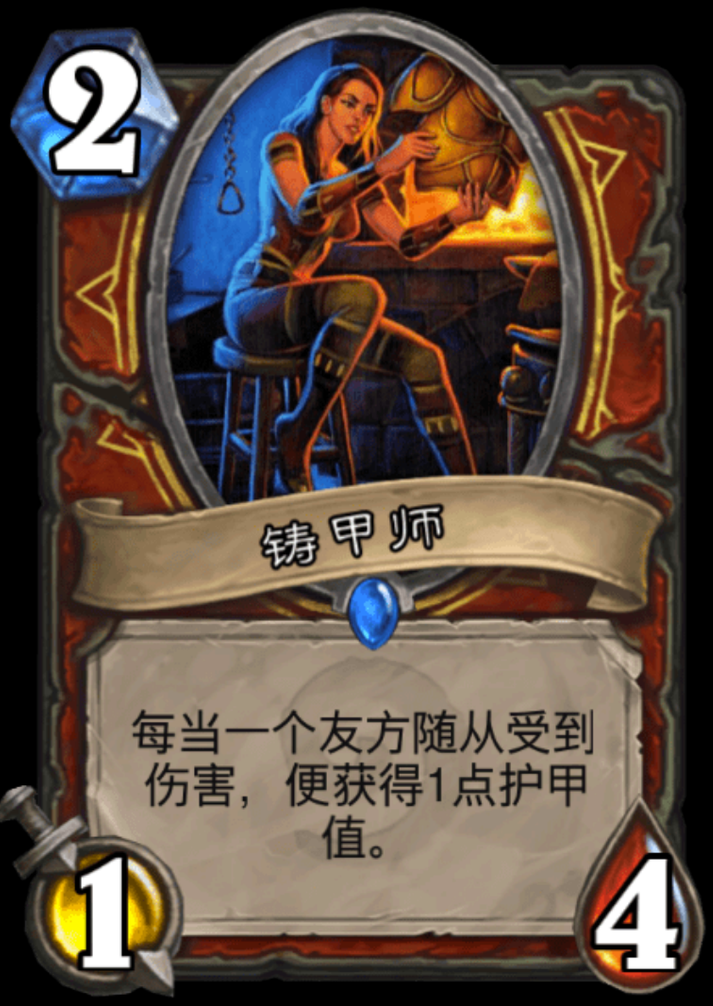
找到因子, 弄很多

登录态

注册用户，模拟登录

各种验证码

手工,OCR,AI,打码服务



混淆 编码 图片化

假装摧毁

特征分析

特征分析

请求标识符，钥匙串
甚至用户行为分析

获得标识符
模拟用户行为

不报错，欺诈，甚至注入

先要发现自己被毒，不断尝试伪装

动态加解密

死磕，其人之道，或


9
瘟疫使者诺斯
每当一个敌方随从死亡，召唤一个1/1的骷髅，并使你的其他随从获得+1/+1。
2
5

瘟疫


沉默暗影
永久**潜行**。隐藏属性值。在你的回合结束时，对所有敌人造成1点伤害。

活着才有希望

混合使用 级别控制


7
先知维伦
使你的法术和英雄技能的伤害和治疗效果翻倍。
7
7


5
侏儒军刀
使一个随从获得冲锋，风怒，圣盾，吸血，剧毒，嘲讽以及潜行。

有趣的反爬者

给进化添点乐

爬虫的道德

道法自然，细水长流

爬虫玩的好，监狱进的早， 数据玩的溜，牢饭吃个够

感谢 与 问答