

# 代理 IP 池

## 起因

在众多的网站防爬措施中，有一种是根据 ip 的访问频率进行限制的，在某段时间内，当某个 ip 的访问量达到一定的阈值时，该 ip 会被拉黑、在一段时间内被禁止访问。

这种时候，可以通过降低爬虫的频率，或者更改 ip 来应对。后者就需要有一个可用的代理 ip 池，以供爬虫工作时切换。

## 思路

- 1、找到一个免费的 ip 代理网站(如：西刺代理)
- 2、爬取 ip (常规爬取 request, 建议 Selenium)
- 3、验证 ip 有效性 (携带爬取到的 ip, 去访问指定的 url, 看返回的状态码是不是 200)
- 4、存储 ip (写到文档或者数据库)

## 代码实现

### 1、封装类

封装 ProxyIpPool 类，实现 ip 的抓取，验证，存储，随机获取代理 ip 功能

初始化 mysql 数据库连接

## 2、已有记录读取

从数据库中读取已有并代理 ip，并检验其有效性

## 3、下载获取

从西刺代理或其他代理 ip 网站下载免费的代理 ip

## 4、有效性检验

检验 ip 是否有效

## 5、存储

代理 ip 存储到 mysql 数据库

## 6、代理接口

随机提取可用的代理 ip

## 7、无效剔除

删除指定的记录

## 8、关闭

关闭浏览器，关闭数据库

## 9、测试

测试代理 ip 池功能

### 代理 IP-API 接口的使用

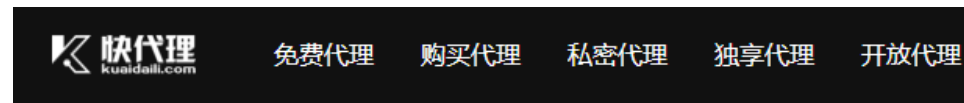
以快代理为例

#### 快代理

<https://www.kuaidaili.com/>



## 会员注册



### 欢迎注册快代理

邮箱

密码

确认密码

QQ号

手机号

验证码  [获取验证码](#)

☐ 我同意《快代理网站服务条款》|《法律声明和隐私权政策》

[创建账户](#)

已有快代理帐号? [直接登录](#)

## 私密代理

### 产品特性

<p><b>IP量大</b> 每天可用私密代理IP10000+ 一次可提取超过350个IP</p>	<p><b>极速访问</b> IP白名单或用户名密码授权使用 访问国内网站响应时间&lt;1秒</p>	<p><b>HTTPS / Socks</b> 支持Https/Socks4/Socks5 <small>新</small> 支持GET/POST/CONNECT请求</p>
<p><b>高匿名</b> 对方网站不知道您使用了代理 完美隐藏本机IP</p>	<p><b>API支持</b> 支持多台设备同时调用API 支持高频率调用</p>	<p><b>多设备支持</b> 多台设备同时使用私密代理 PC和移动设备均可使用</p>

功能丰富，可帮助您完成多项任务



**数据采集**  
爬虫抓取 网页采集



**网购下单**  
淘宝 天猫 京东



**注册和登录**  
登录或注册网站、APP账户



**投票**  
微信投票 网页投票



**挂软件**  
QQ YY 旺旺 ...



**浏览网页**  
浏览国内各大网站

## 购买

计数版新	私密代理1000+	私密代理3000+	私密代理5000+
<b>¥60</b> 起	<b>¥60</b> 1天	<b>¥162</b> 1天	<b>¥240</b> 1天
<a href="#">立即购买</a>	<a href="#">立即购买</a>	<a href="#">立即购买</a>	<a href="#">立即购买</a>
每天累计可用IP量: 1000~12000 实时可用IP量: 350+	每天累计可用IP量: 1000+ 实时可用IP量: 30~35	每天累计可用IP量: 3000+ 实时可用IP量: 90~105	每天累计可用IP量: 5000+ 实时可用IP量: 150~175
新品8折优惠			

## 私密代理 API 接口

具有编程能力的开发者可在程序中调用 api 实时批量获取 IP 列表。

所有私密代理 IP 都是五位随机端口、极速（响应时间<1 秒）、高匿名、支持 HTTP GET/POST/CONNECT 方法、支持 HTTPS

**/api/getdps**

### 接口说明

1. HTTP GET 接口
2. 参数和返回内容均采用 utf-8 编码
3. 接口最快调用频率: 1 秒 10 次

## 私密代理API接口

### /api/getdps

具有编程能力的开发者可在程序中调用api实时批量获取IP列表。

所有私密代理IP都是五位随机端口、极速（响应时间<1秒）、高匿名、

## 接口说明

### 1. HTTP GET接口

2. 参数和返回内容均采用utf-8编码

3. 接口最快调用频率: 1秒10次

### 生成API链接

#### 参数说明

参数	是否必填	参数说明	取值说明
orderid	是	订单号	<a href="#">查看我的订单</a>
num	是	提取数量	例如: 100
pt	否	提取的代理ip类型	1: http代理(默认) 2: socks代理
ut	否	提取的代理ip可连续使用多长时间 根据您一个代理ip需要使用多久来选择筛选条件	0: 不筛选(默认) 1: 可用10分钟 2: 可用30分钟 3: 可用1小时 4: 可用2小时
f_loc	否	提取结果包含地区信息	取值固定为1
f_et	否	提取结果包含此代理从提取时算起的可用时间 (单位: 秒)	取值固定为1
dedup	否	过滤今天提取过的IP 不带此参数代表不过滤	取值固定为1
format	否	接口返回内容的格式	text: 文本格式(默认) json: json格式 xml: xml格式
sep	否	提取结果列表中每个结果的分隔符	1: \n\n分隔(默认) 2: \n分隔 3: 空格分隔 4:  分隔

## 返回值格式

### 文本格式

正常返回代理列表，代理分隔符参见 sep 参数，代理格式 ip:port

格式样例：

```
183.207.226.9:15867
120.197.85.171:23965
120.198.230.55:18923
```

120.198.230.112:21576

118.194.217.134:20523

183.232.29.132:18204

异常返回错误码和错误提示, 格式 ERROR(错误码): 错误提示

格式样例:

ERROR(-1): 参数错误

## json 格式

正常返回格式样例:

```
{
  "msg": "",
  "code": 0,
  "data": {
    "count": 10,
    "dedup_count": 10,
    "order_left_count": 990,
    "proxy_list": [
      "124.172.117.189:19812",
      "219.133.31.120:26947",
      "183.237.194.145:28436",
      "183.62.172.50:23485",
      "163.125.157.243:17503",
      "183.57.42.79:26483",
      "202.103.150.70:17251",
      "182.254.129.124:15395",
      "58.251.132.181:20659",
      "112.95.241.76:21948"
    ]
  }
}
```

### 返回值说明

字段名称	说明
code	返回码。取值：0代表成功 非0代表失败 具体参见 <a href="#">错误码一览表</a>
msg	错误信息
data	包含接口返回的数据
data.proxy_list	返回的代理ip列表
data.count	返回的代理ip数量
data.dedup_count	返回的不重复的代理ip数量(计数版专有)
data.order_left_count	订单剩余提取余额(计数版专有)