

现在越来越多的网站开始注意保护自己的数据，通过各种策略来反抗异常的爬虫。其中最常见的策略是限制客户端IP的访问频率，通常表现为当爬虫客户端访问网站速度过快时，网站会在一段时间内返回403、503之类的错误（例如谷歌、亚马逊）或返回验证码页面（例如大众点评网、去哪儿网）。

而绕过网站访问频率限制这种反采集策略最通用的办法就是用HTTP代理，具体做法：在爬虫内严格控制每个代理IP的两次访问目标网站的间隔的前提下，同时使用多个代理IP并行抓取，以加提高整体采集速度。

需要注意的是，并非所有的HTTP代理都适合于Web数据采集。网上免费的HTTP代理大多是不稳定的透明代理（俗称“野代理”），不能有效的用于Web数据采集。事实上，只有高匿稳定的HTTP代理才能有效的应用于爬虫中，关于这点，在这篇文章里我们做了详细的介绍<http://www.site-digger.com/html/articles/20130114/50.html>。

鲲之鹏目前为广大爬虫客户提供以下几种类型的稳定高匿HTTP代理解决方案：

1. 美国HTTP代理；



- o IP固定不变；
- o 分独享和共享两种；

2. 国内HTTP代理；



- o IP固定不变；
- o 分独享和共享两种；

3. 国外动态IP HTTP代理；



- o IP每隔一段时间变换一次；
- o 共享IP；
- o IP总量超过100W，而且会不断更新；

4. 国内ADSL动态IP HTTP代理；



- 可通过调用重拨API获取新的IP;
- 独享IP;
- 有多个城市可选;

注意：我们不提供VPN服务，我们的HTTP代理服务仅供WEB爬虫使用。



西安鲲之鹏网络信息技术有限公司

选择我们，所有数据都是你的！



公司名称：西安鲲之鹏网络信息技术有限公司

网 址：<http://www.site-digger.com/>

地 址：陕西省西安市雁塔区长安中路99号长安文化综合大厦11821室

邮 编：710061

联系电话：029 - 87553281

手 机：13571845363 齐工

13389148466 周工

客 服 QQ：1649677458 或 312602670

邮 箱：hello@site-digger.com
