



链滴

# 网站为什么要反爬虫?

作者: [TIANQIIP](#)

原文链接: <https://ld246.com/article/1633931032031>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

互联网有很多业务或者说网页，是不需要用户进行登录的，这些不需要登录的页面，往往会包含大量聚合信息，比如新闻门户网站、视频门户网站、搜索引擎，这些信息是公开的，是可以被爬虫抓取的。

## 一、网站为什么要反爬虫？

### 1、爬虫占总PV比例较高，浪费服务器资源

通过程序进行 URL 请求去获得数据的成本是很低的，这就造成大量低质量网页爬虫在网络横行，对标网站产生大量的访问，导致目标网站的服务器资源大量消耗，轻则影响正常用户的访问速度，重则致网站服务不可用。

### 2、公司可免费查询的资源被批量抓走，丧失竞争力

很多软件的价格可以在非登录状态下直接被查询，如果没有反爬虫，竞争对手就可以批量复制网页信，抓取到软件的价格、资源等各种信息，时间长了企业的竞争力就会大大减少。

## 二、我们在反什么样的爬虫？

### 1.黄牛恶意竞争

黄牛党利用恶意爬虫遍历航空公司的低价票，同时批量发起机器请求进行占座，导致航班座位资源被持续占用产生浪费，最终引发航班空座率高对航空公司造成业务损失，并且损害正常用户的利益。

### 2.没人去停止的失控爬虫

互联网将近60%的访问量都是爬虫带来的，网站已经对这些爬虫程序进行了限制，阻止它们爬取数据可这些爬虫依然孜孜不倦地运行着，即使抓取不到任何数据。这是因为一些托管在某些服务器上的爬，已经是无人认领的状态了，没有人停止。

### 3、同行竞争对手

企业需要数据来分析用户行为、自己产品的不足之处以及竞争对手的信息等，就会爬取竞争对手的信，像电商类网站、招聘类网站就会爬取竞争对手的产品信息，为保证自己的产品竞争力，企业往往会对此类爬虫。

### 4、网站点击欺诈

投放广告通常是为了触达符合网站定位的潜在消费者，而恶意爬虫造成的点击欺诈使得广告的点击率高，使得网站承担了本不应承担的点击费用，给网站造成实实在在的利益损失。