## 爬虫和法律

### 爬虫背后的法律风险

个人信息是爬虫的一个禁区，通常我们都知道，作为爬虫工程师不应该爬取个人信息。那么在我国法律中，关于个人信息的描述如下：

**《中华人民共和国网络安全法》**

2016年11月7日发布的《中华人民共和国网络安全法》明确“个人信息”是指以电子或者其他方式记录的能够单独或者与其他信息结合识别自然人个人身份的各种信息，包括但不限于自然人的姓名、出生日期、身份证件号码、个人生物识别信息、住址、电话号码等，就网络服务中的个人信息保护问题作出系统规定如下：

1、保护个人信息：

网络产品、服务具有收集用户信息功能的，其提供者应当向用户明示并取得同意；涉及用户个人信息的，还应当遵守本法和有关法律、行政法规关于个人信息保护的规定。

2、合法收集使用个人信息：

网络运营者收集、使用个人信息，应当遵循合法、正当、必要的原则，公开收集、使用规则，明示收集、使用信息的目的、方式和范围，并经被收集者同意。不得收集与其提供的服务无关的个人信息，不得违反法律、行政法规的规定和双方的约定收集、使用个人信息，并应当依照法律、行政法规的规定和与用户的约定，处理其保存的个人信息。

3、不得向他人提供个人信息：

网络运营者不得泄露、篡改、毁损其收集的个人信息，未经被收集者同意，不得向他人提供个人信息，但经过处理无法识别特定个人且不能复原的除外。

**《规范互联网信息服务市场秩序若干规定》**

1、搜集须经许可：未经用户同意，不得搜集与用户相关、能够单独或者与其他信息结合识别用户的信息（“用户个人信息”），但法律法规另有规定除外；

2、限定搜集范围和用途：经用户同意搜集用户个人信息的，应当明确告知搜集和处理用户个人信息的方式、内容和用途，不得收集其提供服务所必需以外的信息，不得将用户个人信息用于其提供服务之外的目的；

3、用户个人信息保障：互联网信息服务提供者应当加强系统安全防护，妥善保管用户个人信息，未经用户同意，不得向他人提供用户上载信息，但是法律法规另有规定的除外。

### 网络爬虫怎么做比较好

1、如果你爬虫获取信息，是为了证明被爬公司的数据造假，其获取的数据也都是通过公开渠道可以查询的，那么，在获取信息后公布于众的行为并不违法，也不侵犯被爬公司的民事权益。但是，如果你爬取公开免费的信息，是用来进行违法操作，比如造假、诽谤等，就有隐患了。

2、如果你利用爬虫获取其他公司的公开信息数据，用于自身公司的经营。而被爬公司的信息是投入了大量人力、财力，经过常年积累获得的，并且被爬公司本身也采取了反爬措施。这种情况下，虽然信息是公开的，但信息本身具有较高的商业价值，能够给使用者带来商业利益，此时的爬虫也是违法的。

3、如果你未经平台授权，强行突破反爬措施，导致被爬网站的运行受到严重影响，这种行为明显是违法的，这里涉及的就是我们上一个技术篇讲到的破解犯罪了。

4、如果你是第三方应用，想要通过开放平台获取用户信息时，更要注意授权问题。从用户对平台的授权、平台对第三方的授权、再到用户对第三方的授权，三重关卡都要通过才可以