### ****1.Robots协议****

网络爬虫排除协议（Robots Exclusion Protocol）

这目前是一个道德规范，目前还未了解到有相关的法律规范

### ****2.网络爬虫的约束****

有名的两个案子

百度大战360：<https://www.williamlong.info/archives/3204.html>

爱帮网大战大众点评：<https://yq.aliyun.com/articles/483576>

我们在用爬虫的时候一定要遵循规范，**不然会给自己和他人都带来麻烦**

### ****3.Python爬虫的流程****

其实Python写爬虫非常的简单，就三步：获取网页，解析网页，存数据

## ****4.三个步骤的技术要求****

**1、获取网页**

基本要求：request，urllib和selenium

进阶打怪：多线程多进程抓取，登录抓取，突破IP封禁和服务器抓取

**2、解析网页**

基本要求：re正则表达式，BeautifulSoup和lxml

进阶打怪：解决中文乱码

**3、存储数据**

基本要求：存入的txt和csv文件

进阶打怪：存到数据库去，就想MySQL和MongoDB等等