## 为什么要做爬虫

首先请问：都说现在是"大数据时代"，那数据从何而来？

1)企业产生的用户数据：

[百度指数](http://index.baidu.com/)：<http://index.baidu.com/>

2)数据平台购买数据：

[贵阳大数据交易所](http://trade.gbdex.com/trade.web/index.jsp)：<http://trade.gbdex.com/trade.web/index.jsp>

3)政府/机构公开的数据：

[中华人民共和国国家统计局数据](http://data.stats.gov.cn/index.htm)：<http://data.stats.gov.cn/index.htm>

4)数据管理咨询公司：

[麦肯锡](http://www.mckinsey.com.cn/)：<http://www.mckinsey.com.cn/>

5)爬取网络数据：

如果需要的数据市场上没有，或者不愿意购买，那么可以选择招/做一名爬虫工程师，自己动手丰衣足食。

## 二、爬虫是什么

网络爬虫是一种按照一定的规则，自动地抓取万维网信息的程序或者脚本。

简单地说：就是用事先写好的程序去抓取网络上所需的数据，这样的程序就叫网络爬虫。

Urllib和requests模块是发起http请求最常见的模块。

**安装：**

**pip install requests**