### 数据建模交接

1. 抓包工具Charles的下载和安装

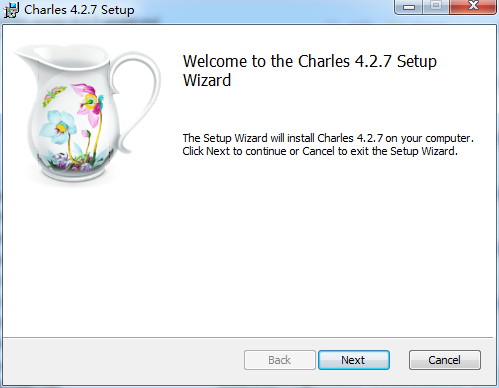
下载网址：<https://www.charlesproxy.com/download/>

生成注册码网址：https://www.zzzmode.com/mytools/charles/

安装教程：

Windows版

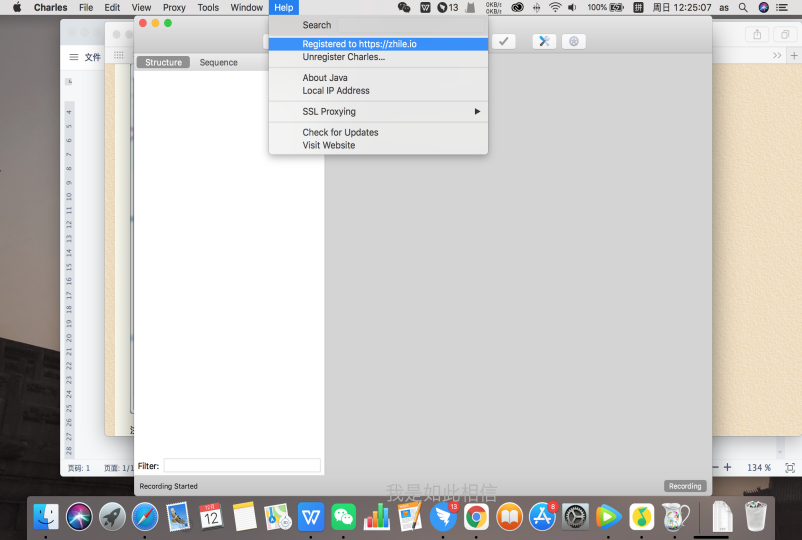
一路按下next即可安装（保证电脑里没有其他charles即可）



Mac版

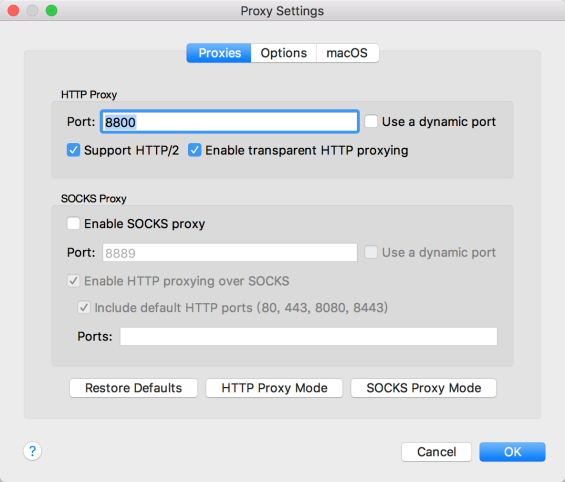
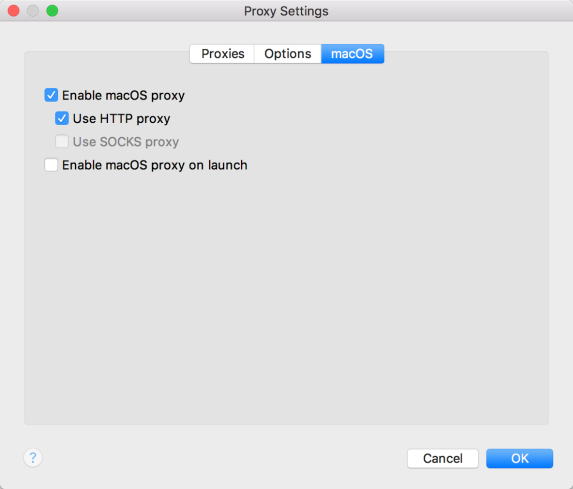
正常安装

安装好之后，打开Charles，点击help-register，输入注册码。



Charles设置

点开proxy-proxy setting，设置端口号（port），enable macOS proxy launch是用于抓取本台电脑上的URL，在抓取手机端的时候可以不勾选。

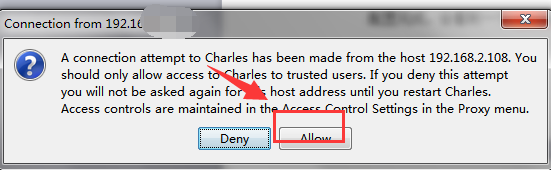
 

手机端设置

点开手机：设置-无线局域网，连接Wi-Fi后，对Wi-Fi进行设置，服务器ip要与电脑端IP地址一致，端口号要和Charles中设置的一致。

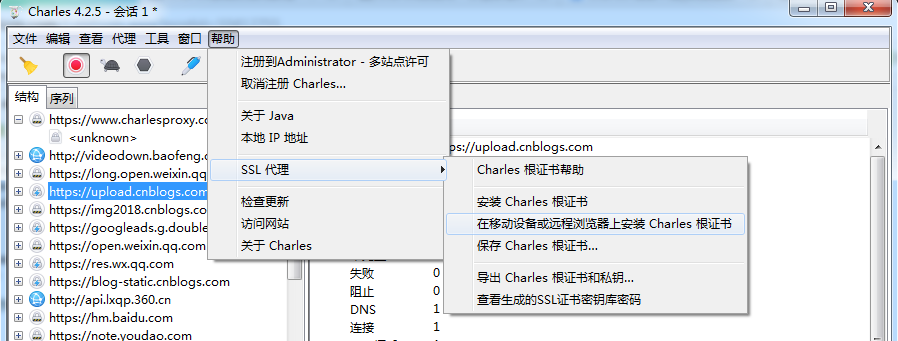


配置完，Charles会跳出弹窗，点allow。



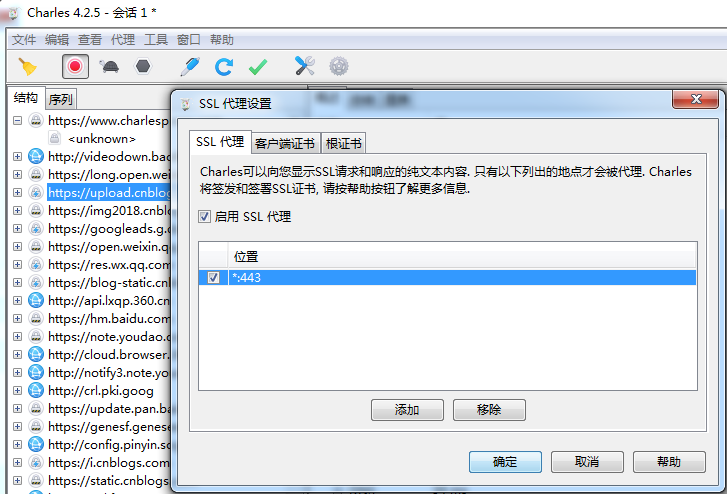
以上步骤完成，便可以抓去http形式的URL，并能获取其主域名下的子域名。但要获得https形式下的子路径的URL，还需要在电脑端和手机端下载证书。

电脑：



手机：在设备上打开这个网址 http://www.charlesproxy.com/getssl 安装 Charles SSL 证书。Ps：ios手机，需要信任该证书。

相关配置：选择Proxy | Proxy Settings，弹出proxy设置选项卡，勾选Enable SSL Proxying，在Location部份选择add，按如下图添加，抓取任意站点、443端口的数据。



到这，抓包工具准备完毕。

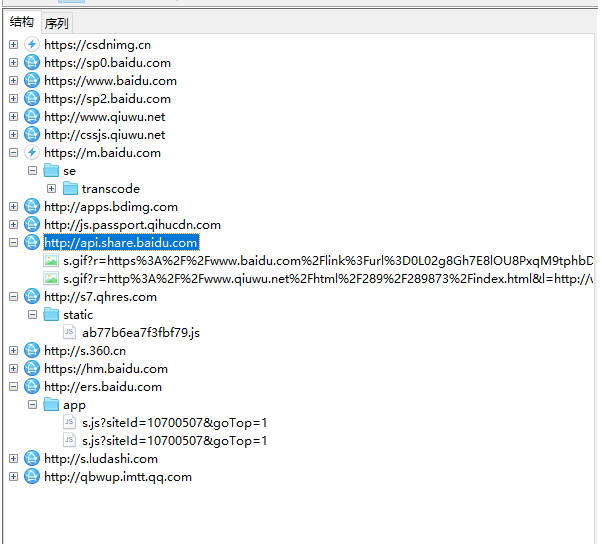
Charles基本功能介绍：



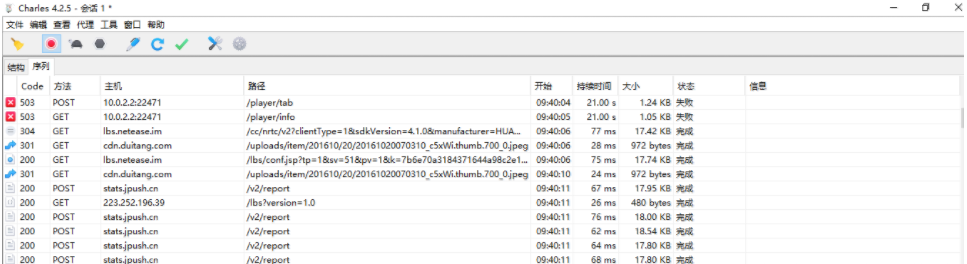
从左到右分别为：

（1）清除当前会话  
（2）停止记录  
（3）开始限流/停止限流  
（4）开始断点/停止断点  
（5）根据所写内容撰写新请求  
（6）重发选定请求  
（7）工具  
（8）设置

显示模式：有结构和序列两种模式  
（1）结构模式：可以很清晰的看到请求的数据结构，而且是以域名划分请求信息的，可以很清晰的去分析和处理数据。



（2）序列：可以看到全部请求，这里的结果以数据请求的顺序来显示，可以根据时间、大小等自由排序



过滤设置

方法一：软件下方的过滤可以进行简单的过滤，在主界面的中部的Filter栏中填入需要过滤出来的关键字。例如我们的服务器的地址是：[www.baidu.com](https://link.jianshu.com/?t=http://yuantiku.com/" \t "/Users/as/Documents\\x/_blank),那么只需要在 Filter 栏中填入[baidu](https://link.jianshu.com/?t=http://yuantiku.com/" \t "/Users/as/Documents\\x/_blank)即可。

1. 数据来源介绍

App/网站：这是目标URL的主要来源，如果能找到行业特有的app，抓到的URL就越精准。缺点是这类URL一般比较难分辨，因为app的发开商旗下可能会有很多种类不同的app，涵盖业务广泛，但从有的URL中只能获得开发商的信息，以分辨是哪款app。

微信/抖音/淘宝/快手等的头像：这类app一般加密性较强，只能抓到一些公众号、大v的头像，这类URL一般出数量不大。

广告落地页：这类落地页精准度较高，但大多是https形式的，很难取到子路径。

1. URL校验网站

下面给出两个校验URL准确性的网站，可以用于校验在Charles中抓到的URL的准确性。

站长之家：<http://seo.chinaz.com>

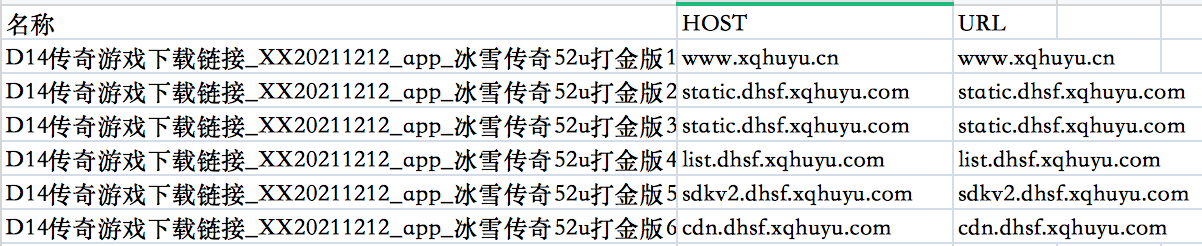
唯凡科技：<https://www.xbeian.com>

在抓完URL后，可以用verification.py脚本检验出一些出数量较大的URL，重点关注其准确性。

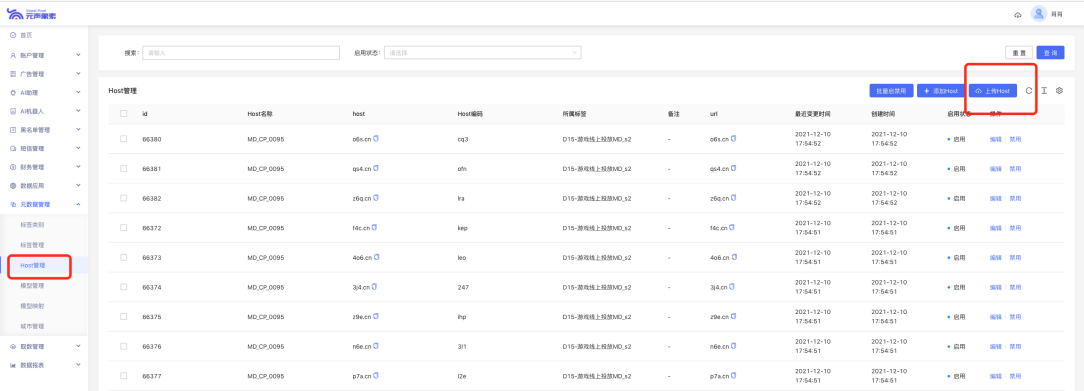
1. 上传平台

将抓到的URL按以下格式调整好，保存为csv.格式，再上传至运营平台。





运营平台的上传



上传完成后，要绑定在特定的标签下。

注：

1. 在抓取过程中，会遇到一些带有baidu、360、weibo等日浏览量很大，人群较杂的主域名，这类URL要小心。
2. 每天建模前要确保手机网络代理重的IP地址与电脑端的一致，且手机和电脑处于同一网络中。
3. 抓取的包全部出现unknow的解决方法

代理-SSL代理-启用SSL代理-添加 如图主机和端口都添加\*。继续：代理-SSL代理-启用SSL代理-添加 如图主机添加\*，端口添加443.

如图所示即可完成 解决unknow问题。如果个别情况出现unknow，可能是本身抓包软件存在个别漏抓问题。可以忽略。

