# 自动化抓取抖音app数据

## 一、准备工作

### 安装mitmproxy

直接在命令行下输入pip install mitmproxy

如果是windows系统，在命令行下输入mitmdump,若无报错信息，则安装成功。

若要让mitmproxy能够监听到手机，还需要对手机进行配置：

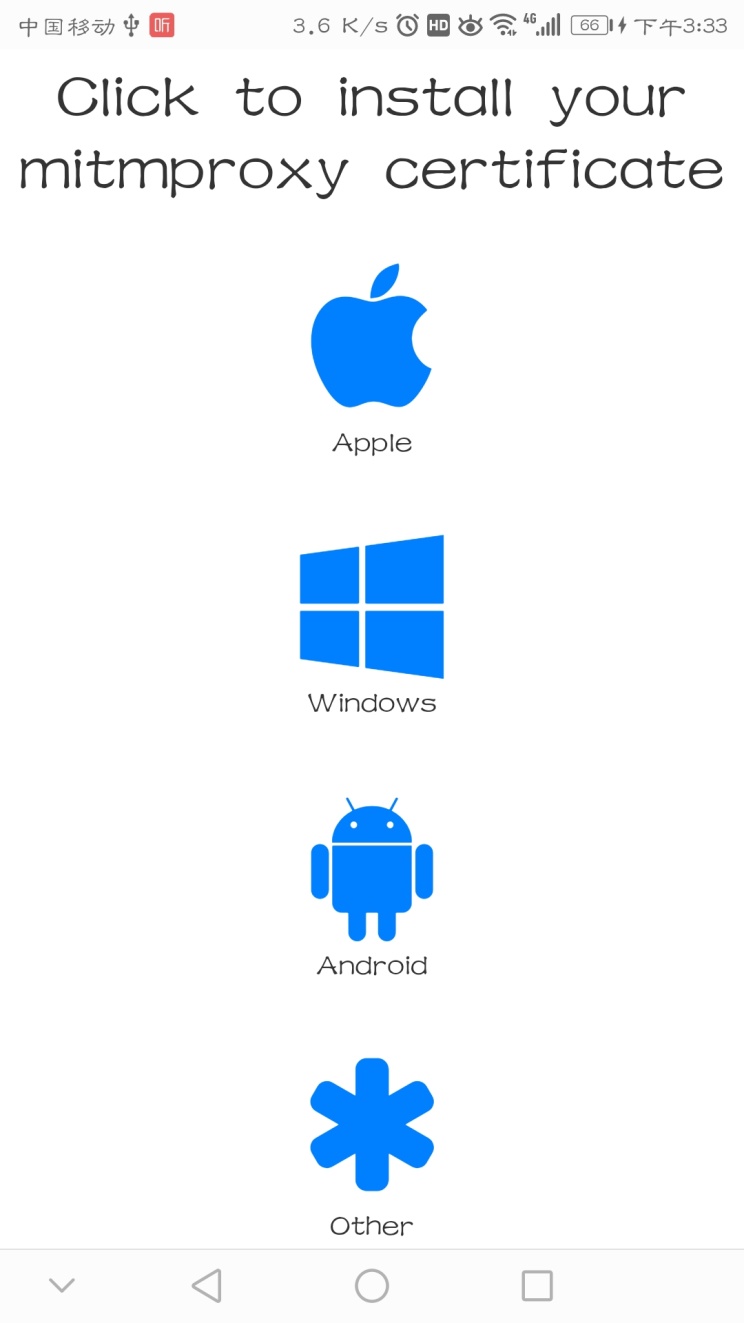
使手机与电脑处于同一局域网下，即在同一wifi下。修改wifi并设置代理。



将ip地址换成自己电脑的ip,点击保存。

配置证书：

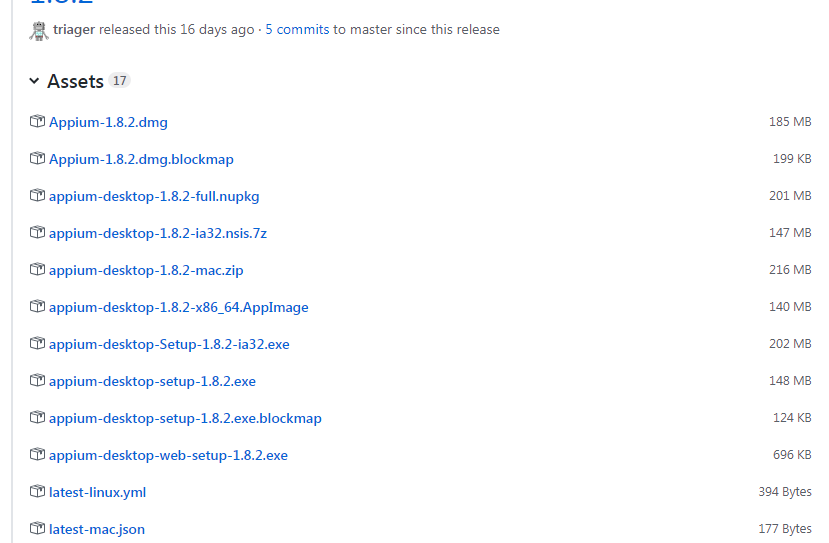
打开<http://mitm.it/>



选择对应的设备，下载安装证书。若显示不是以上图片，请检查服务是否开启，或代理配置是否有问题。

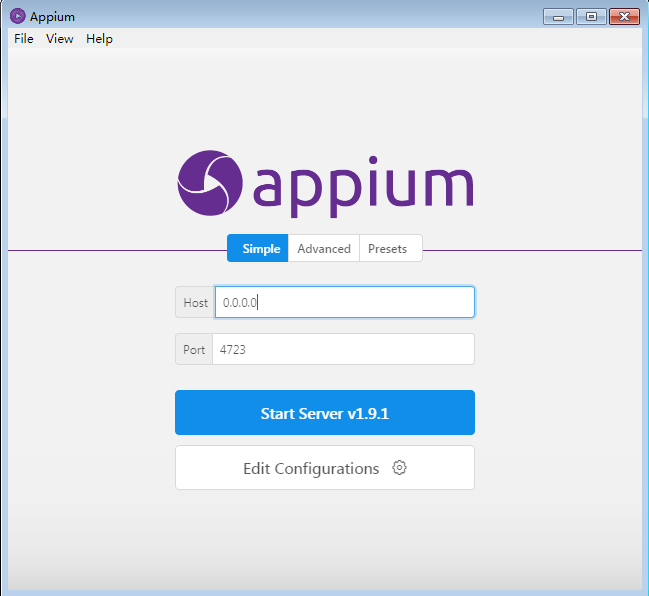
### 安装Appium Desktop

下载地址：<https://github.com/appium/appium-desktop/releases>。



windows平台可下载exe安装包，appium-desktop-Setup-1.8.2.exe

安装成功的界面如下：



### 配置Android开发环境:

为了使appium能够操作手机还需要配置Android SDK。这里推荐安装Android Studio,下载地址：

[https://developer.android.com/studio/index.html?hl=zh-cn](https://developer.android.com/studio/index.html?hl=zh-cn 下载完成后打卡设置页面，选择)

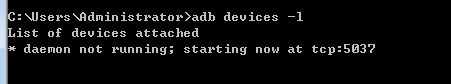
[下载完成后打卡设置页面，选择](https://developer.android.com/studio/index.html?hl=zh-cn 下载完成后打卡设置页面，选择) Android SDK 勾选需要下载的SDK.

sdk下载完成后，在命令行输入adb,若弹出配置信息则证明安装成功。

# 二、小试牛刀

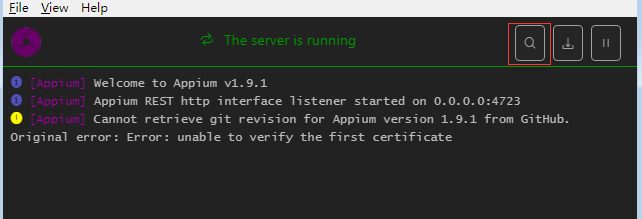
准备好环境了之后，手机连接电脑，打开usb调试功能。

在命令行输入 adb devices –l

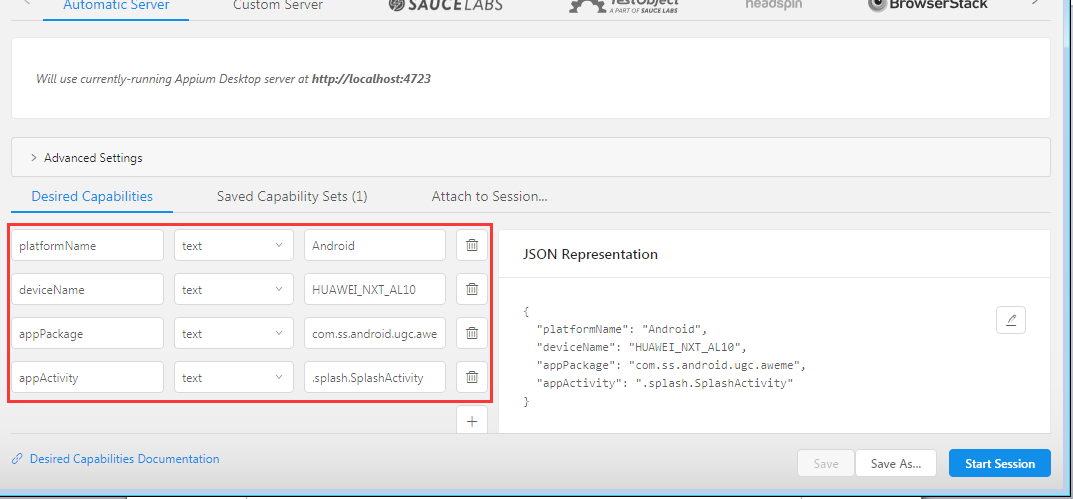


显示该信息则证明PC已连接手机。

之后打开appium客户端，点击start server,再点击右上的放大镜图标



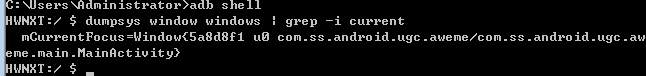
之后，填入以下4个字段信息：



platformName : 设备平台，如安卓、苹果

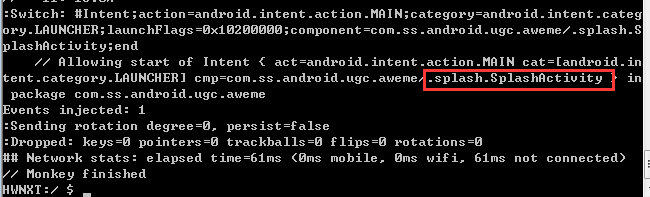
deviceName ： 设备名称

appPackage ： 要打开的app的包名。获取方式：在命令行输入 adb shell,打开手机app,最后输入dumpsys window windows | grep -i current



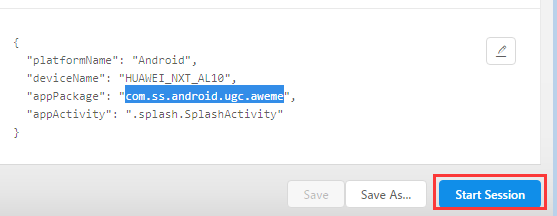
以u0开始/结束的一段字符就是包名。

appActivity ：app的入口Actvity。获取方式：在adb shell中输入monkey -p 包名 -v -v -v 1

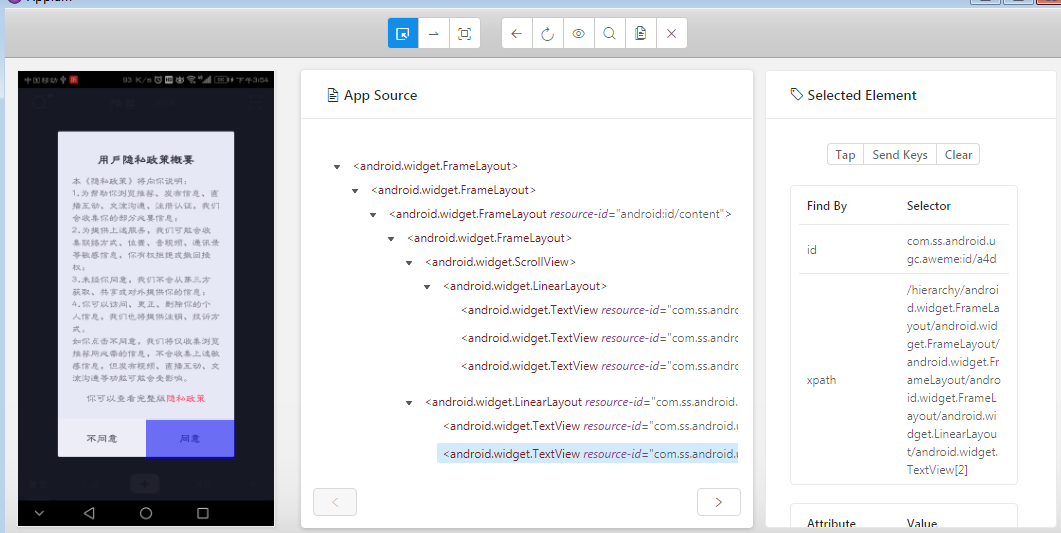


这段就是入口Activity。

以上四个字段配置好后就可以打开app了，点击start session按钮。

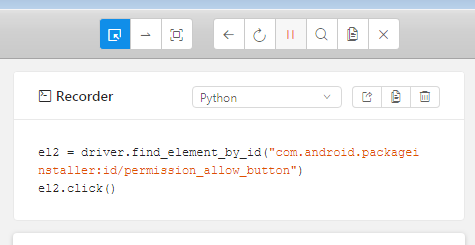


界面如下：



此时鼠标可以点击左侧的组件，中间部分显示了该组件的位置以及id，右侧则显示了该组件的详细信息。选中组件后，可以点击右侧的tap按钮来模拟点击事件。最上层的一个眼睛图标可实现操作的录制，并生成对应代码。

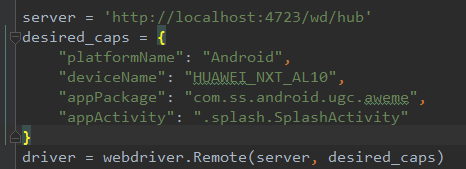
如：当我们选中一个按钮并点击Tap之后



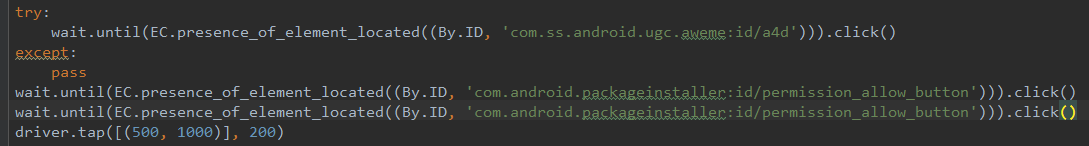
将会生成对应的代码！

下面我将使用代码来操作手机：

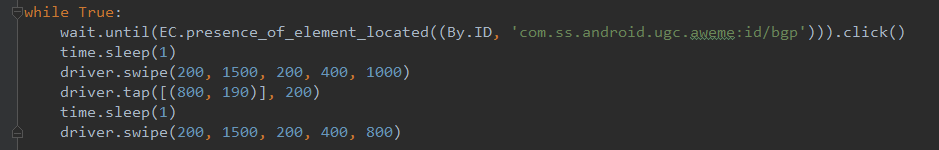
首先打开appium再编写python代码



这段代码就可以打开抖音app了。



这段是点击同意手机的权限以及同意抖音协议。

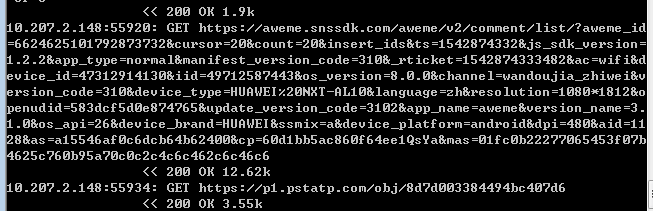


接下来的就是点击评论按钮，滑动评论，关闭评论，滑动下一条视频了。你可能注意到了这里有很多数字，这些数字都是手机屏幕上的坐标，swipe(x1,y1,x2,y2,t)表示在t时间内由x1,y1划到x2,y2。tap则表示点击一个点，最后一个参数表示点击的时长。

点击运行即可操作手机

现在我们以及能够实现自动化了，但是还要实现数据的抓取该怎么办呢，这就需要用到mitmproxy了。

windows 下输入mitmdump，打开抖音app的评论。



我们看到，当滑动评论时，一直在刷新该url,我们可以获取该url下的数据进行查看，

编写一个脚本将数据保存。

# -\*- coding:utf-8 -\*-

from mitmproxy import ctx

import json

def response(flow):

url = 'aweme/v2/comment/list'

if url in flow.request.url:

with open('d://msg.txt','a', encoding='utf-8') as f:

text = flow.response.text

for comment in comments:

f.write(text)

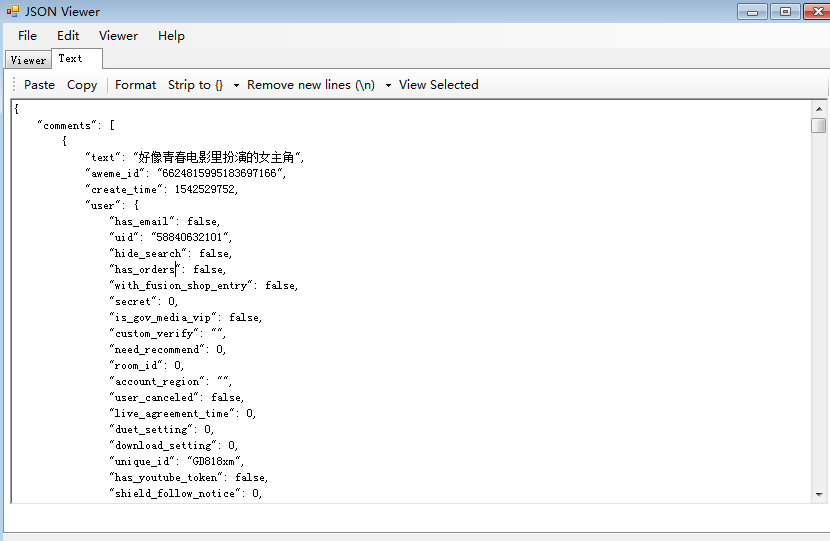
将其命名为script.py, 这样就可以抓取url含aweme/v2/comment/list字符串的文本内容了，现在我们在命令行中进入到该脚本的目录下并输入 mitmdump –s script.py

随后打开评论

这样我们就可以看到抓取到的json数据了：

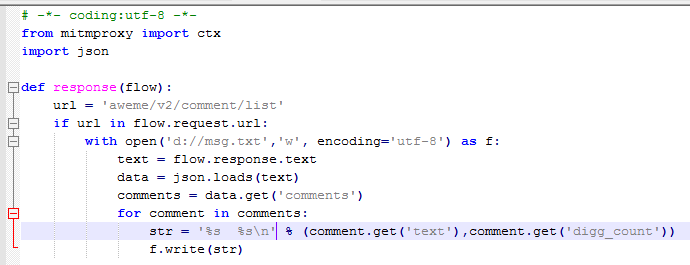


使用工具格式化一下：

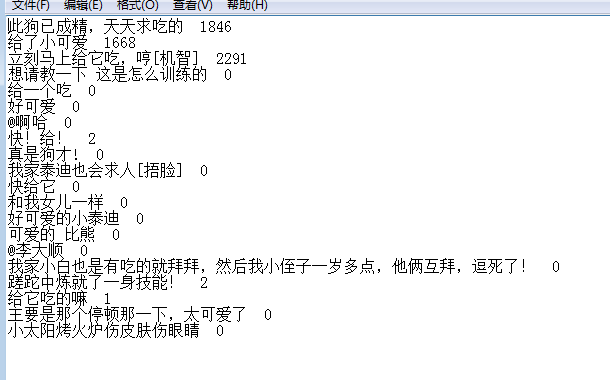


可以清晰的看到内容，id,点赞等。

接下来重新编写script.py的内容，以获取评论和点赞数。



现在，我们使用appium自动操作手机，用mitmproxy获取数据，来让我们试试吧！



成功抓到数据！