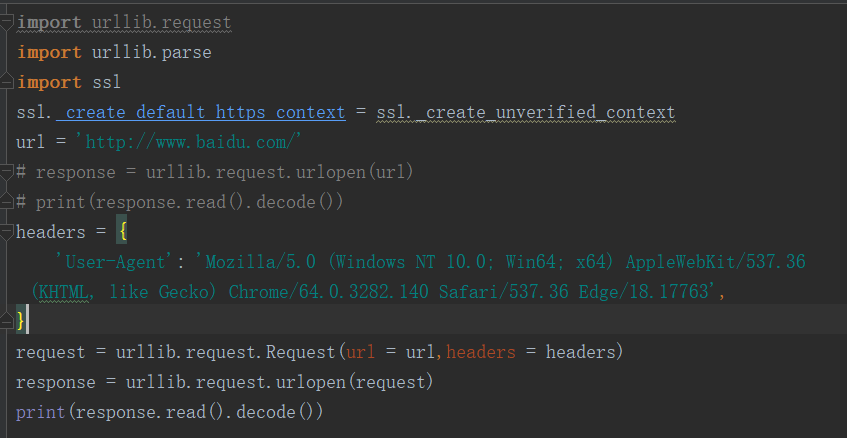
**url**

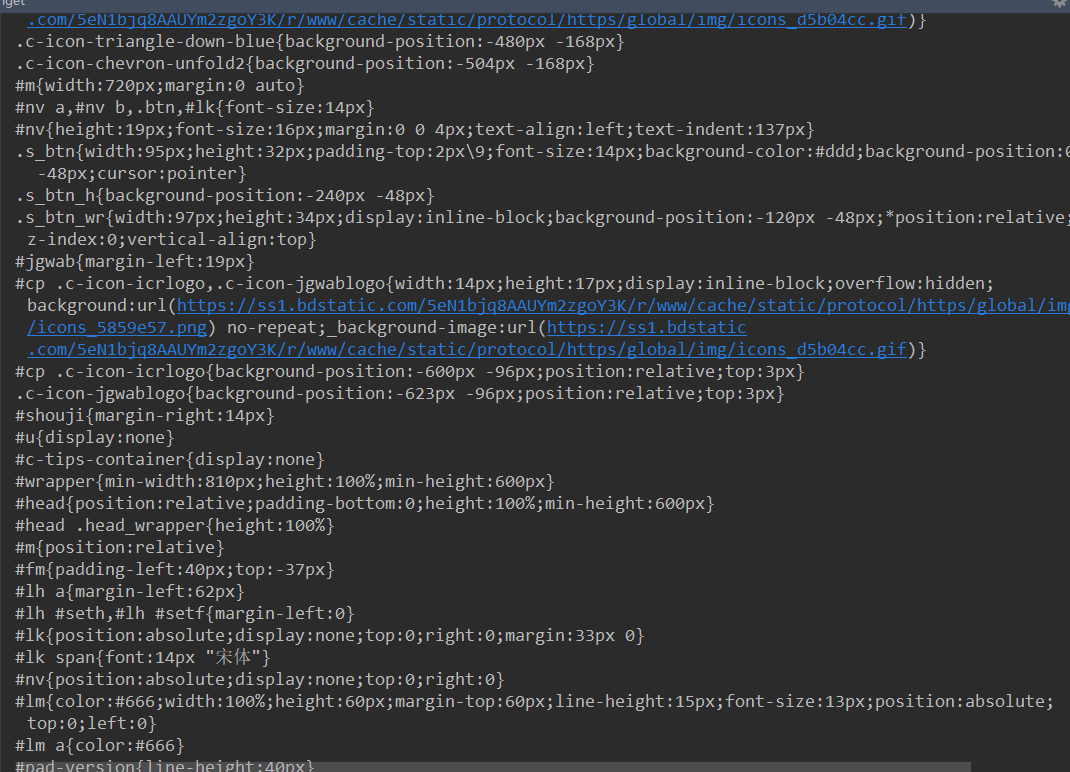
在这里系统的学习url，url就是以字符串形式来描述一个网站的地址，一个url唯一标识一个网页资源，我们所熟悉的网站就是一个url，每个网站都是一个固定且唯一的，我们在爬虫中，每一个爬虫代码都需要url的参数，因为只有我们向这个网站发送请求才能获取内容，得到我们想要的数据。

那么好，我们先来看看代码，简单的练习



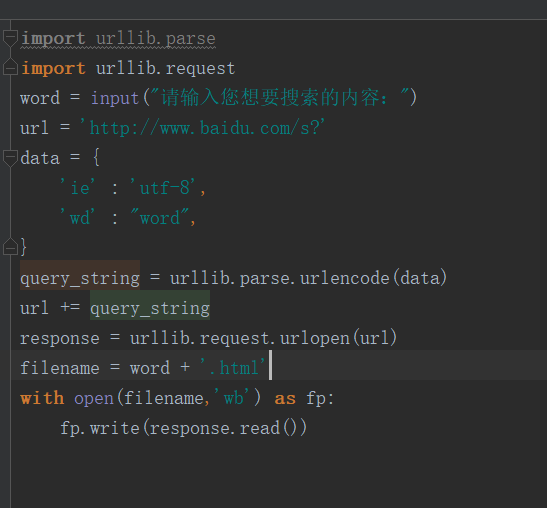
分析代码：

* 首先我们就要写上url地址，也就是这个网站的地址
* 然后之前我们说了，每一个请求我们都需要写一个请求头，因为我们要伪装自己，否则系统判定我们是程序是不会让我们爬取的
* 然后用我们写的url和请求头，向目标网站发送请求，我们得到响应
* 这些响应就是我们想要的内容，然后我们打印出来

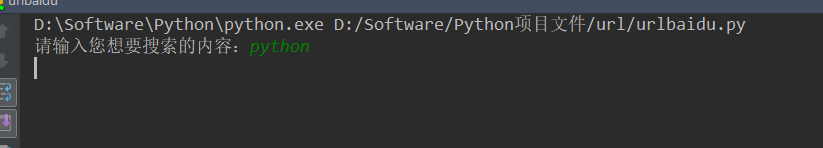
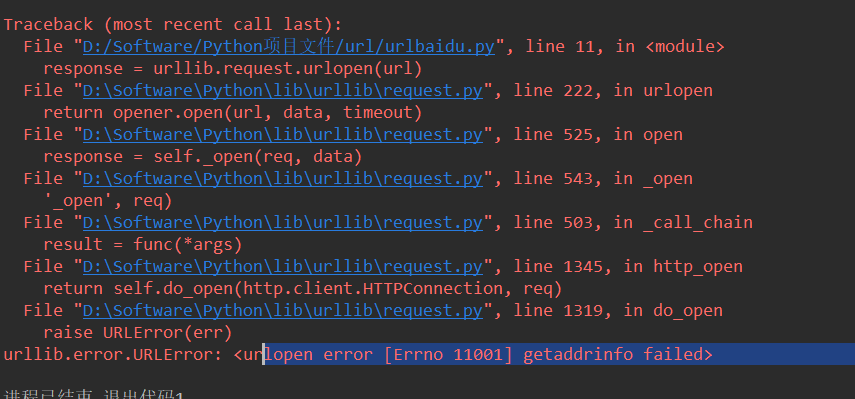
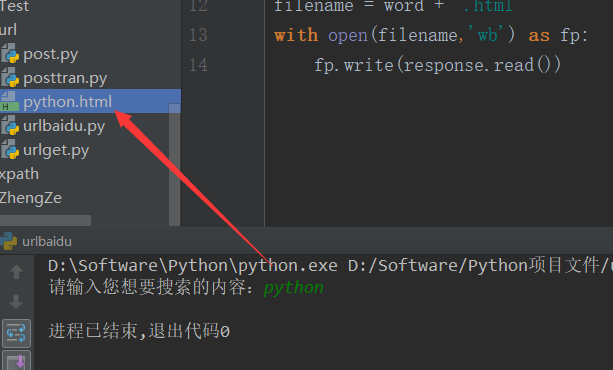
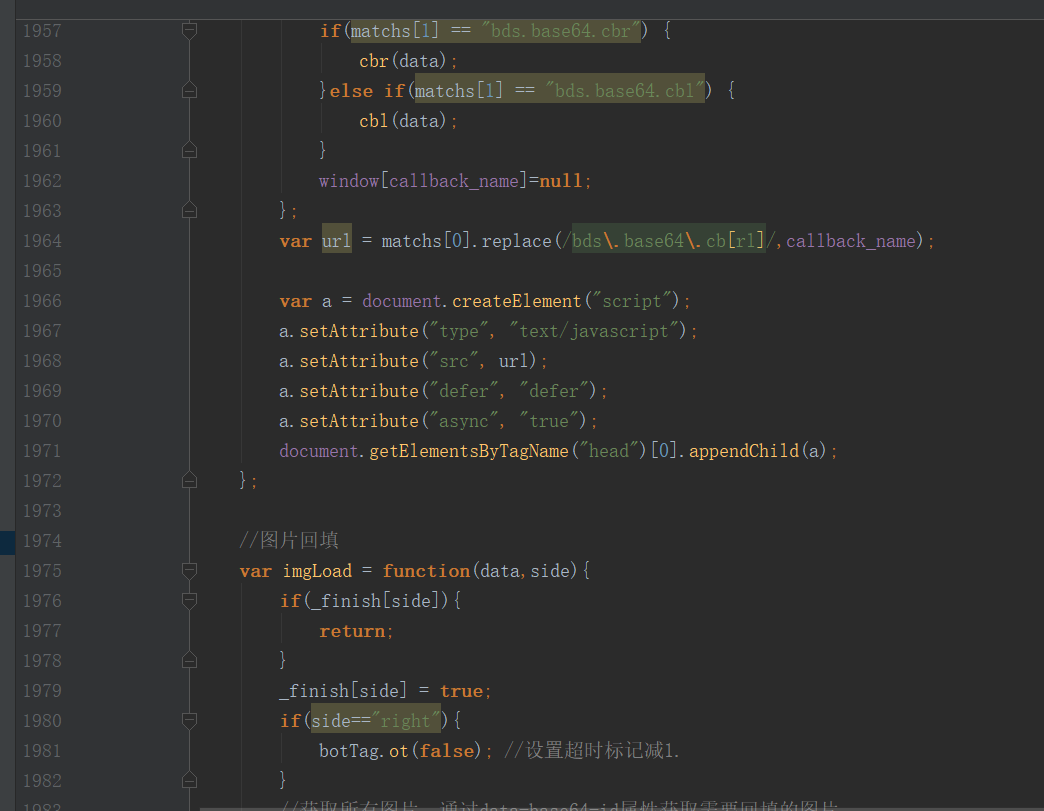


这些就是百度的响应内容

才获取这些内容不爽是不是，百度是用来查内容的，我们连内容都获取不到，说什么爬虫，好，我们就来搜索内容，每一个url里面的内容都代表一个东西，通常有我们需要爬取的内容，页数之类的，我们来看百度的url，里面的ie就是编码格式，我们通用的就是utf8，所以抽取出来，一会写上，那么这个wd呢，就是我们想要爬取的内容，那么好了，我们看代码：

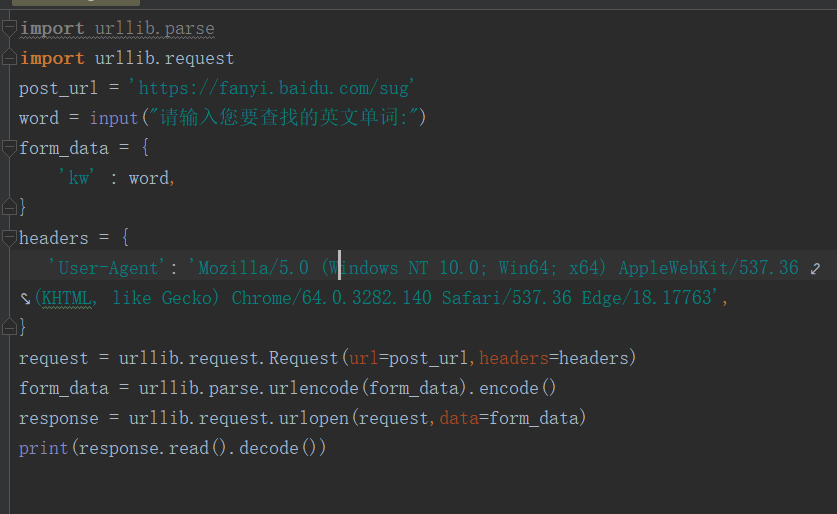


分析代码：

* 首先我们已经分析了，wd是我们要爬取的内容，那么我们定义一个字典，用来存储，存储什么呢，就是我们自己手动输入的内容，然后和url拼接，向这个url发送请求，最后保存到本地
* 我们来看运行过程：
* 
* 这里就需要我们手动输入内容
* 
* 哦，这里报了一个错，查了百度，原来是url不能用单引号，之前犯过这个错误，已经都习惯单引号了，我们要用双引号，修改错误之后，再次运行
* 
* 看我们已经获取到了内容，打开看看
* 
* 我们已经成功获取到了内容

一个不爽，我们再来爬取百度翻译的内容

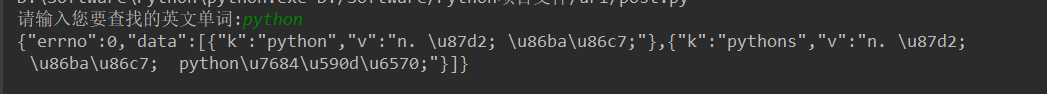
先看代码：



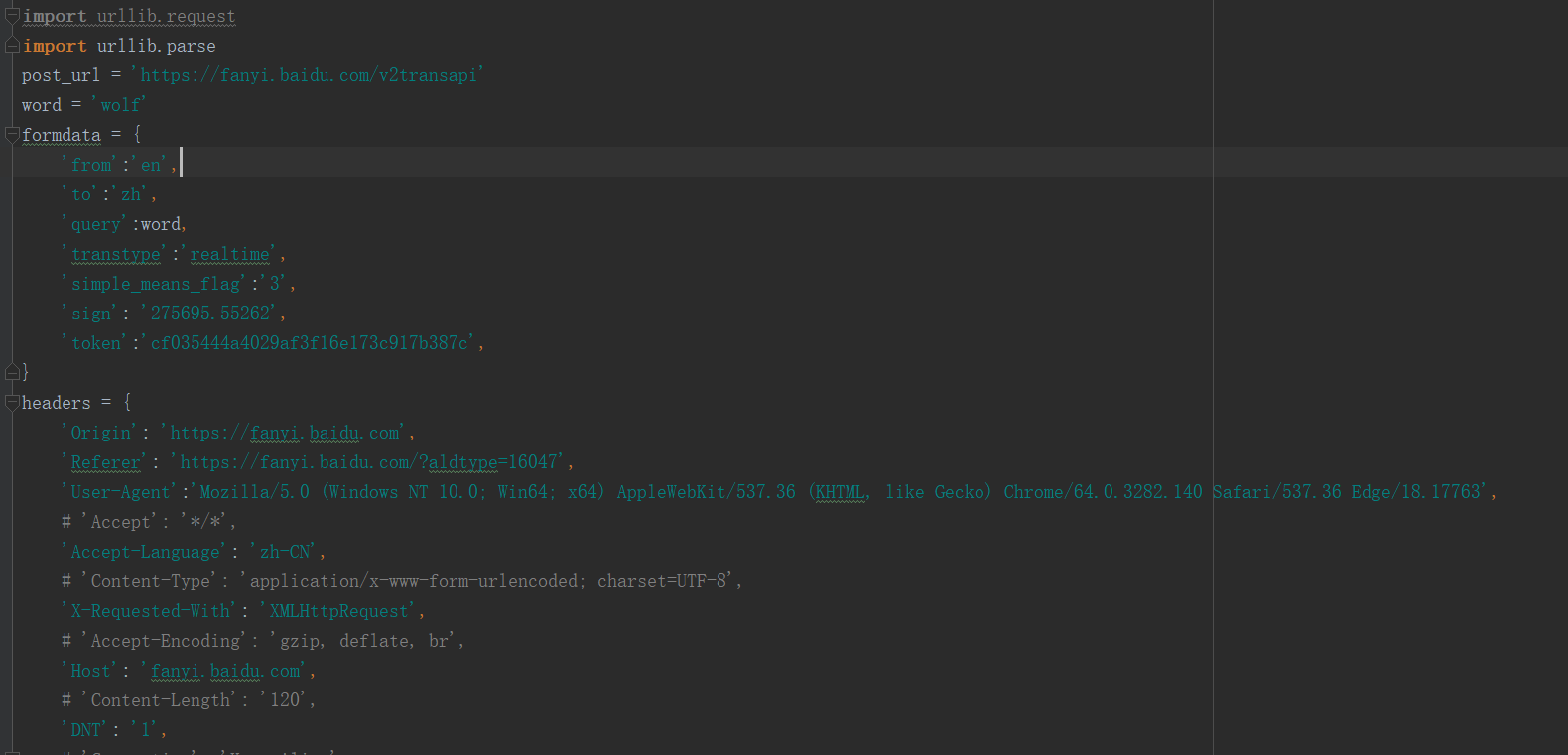
分析代码：

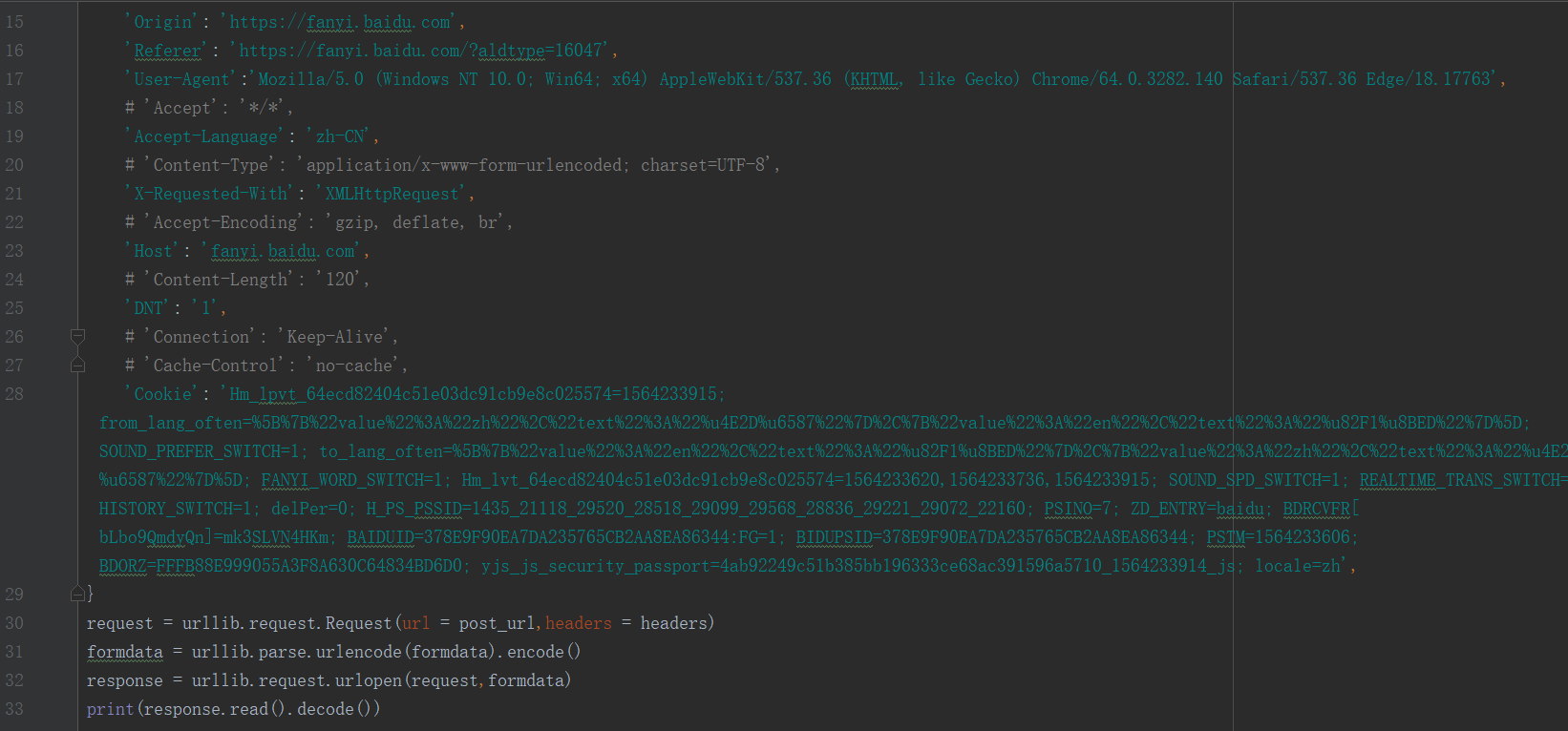
* 首先最重要的就是写入url地址，因为我们要向这个url发送请求，然后我们分析这个url
* 我们可以看到，kw就是我们每次查询的单词，那我们单独提取出来，让我们用户手动输入，那就可以随便查询了，然后再拼接url，向这个url地址发送请求
* 发送请求，获取响应，做了一百八十遍的东西了

我们来看运行结果：

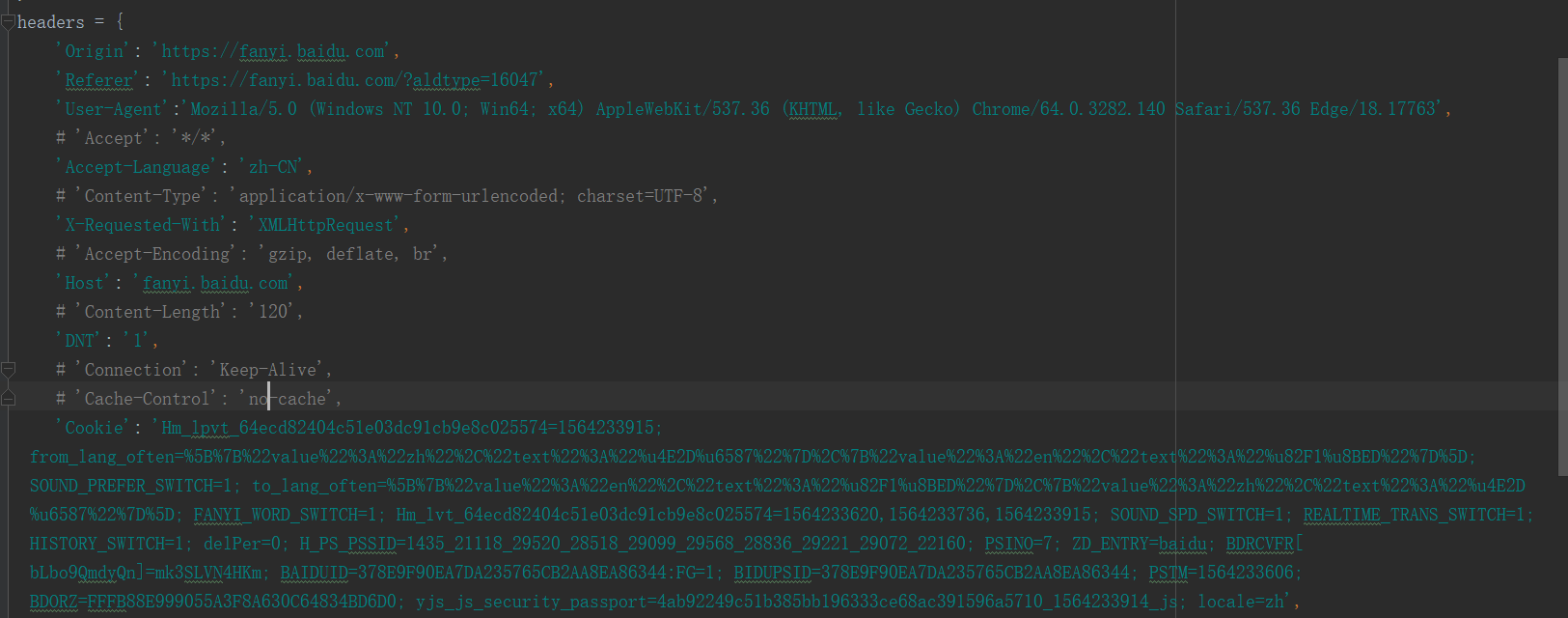


完美

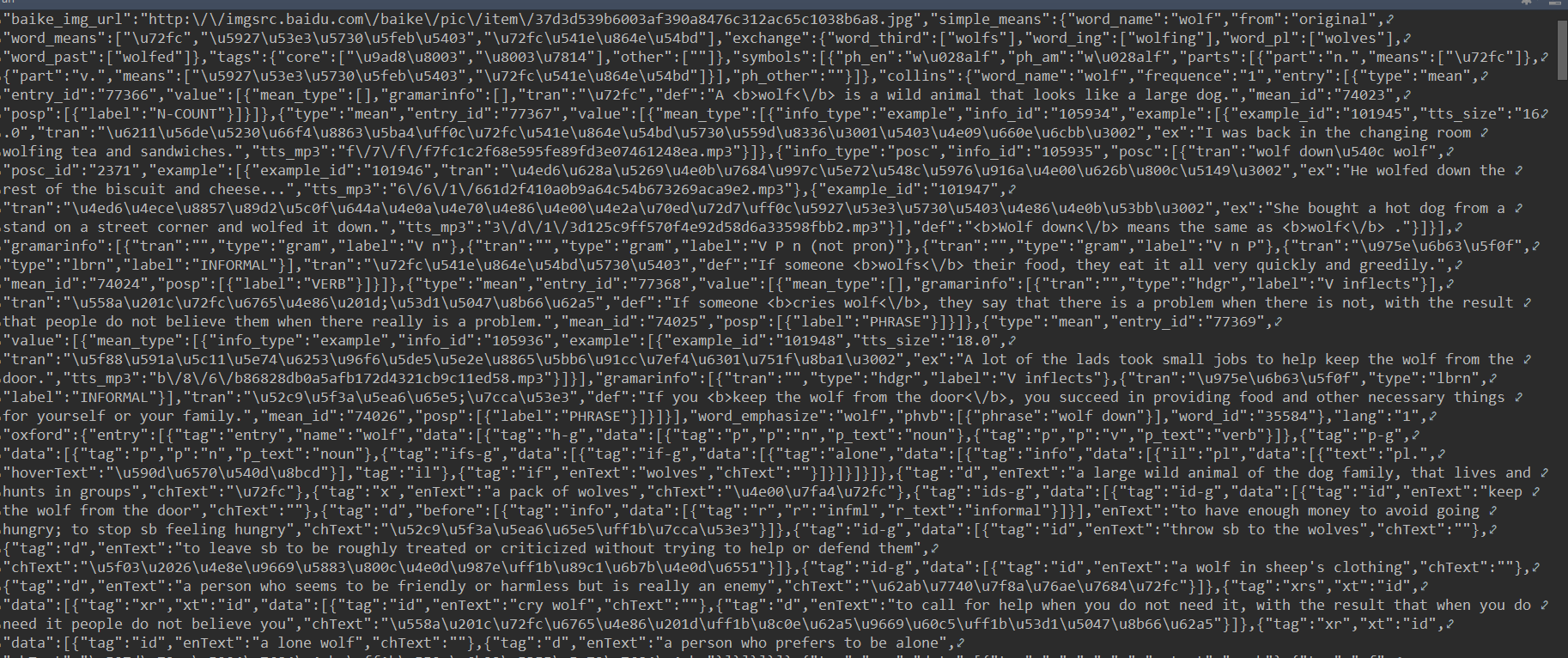
但是后来我想了想，获取的内容还是少啊，我百度翻译看了看，下面还有好多内容，怎么获取呢，我想到了抓包，咱也不知道啊，抓着试试呗，抓完之后，果然在json文件里面有好多内容，我们该怎么获取呢，我想到了cookies，里面是整个页面所有的信息，那么我们看看这里的内容，果然也在，那就开干，看代码：  




分析代码：

* 我们首先看cookies里面的内容，里面有from，to，query，transtype等等内容，这些基本都是固定的，唯一不固定的，是query，这个就是我们要翻译的内容
* 那么好了，先写上url，query还是用word代替，需要我们用户手动输入
* 接下来我们定义字典，用来存储cookies里面的内容，然后query用word代替
* 然后就是老规矩定义请求头，这里的请求头不一样，因为我们用cookies里面的内容，所以我们要这么写
* 
* 把整个cookies里面的内容全部发送过去
* 然后就是发送请求，获取响应，这里还要转化编码格式
* 然后打印内容

我们来看看运行结果：



获取的内容太多了，截图都截不过来了，总之

大功告成！