**静态网页**爬虫的网站是<http://www.zuihaodaxue.com/zuihaodaxuepaiming2016.html>

网页内容不是动态加载

需要用到以下的库：

import requests

from bs4 import BeautifulSoup

import bs4

import pymysql

通过beautifulSoup对获取到的网页源代码进行HTML.praser解析

然后通过节点tree遍历之后找到想要爬取的节点内容，

soup = BeautifulSoup(html, "html.parser")

for tr in soup.find('tbody').children:

if isinstance(tr, bs4.element.Tag):

tds = tr('td')

ulist.append([tds[0].string, tds[1].string, tds[2].string,tds[3].string])

然后存入mysql数据库

由于不知道什么原因这个网站突然访问不了，无法截图，抱歉老师。

**动态网页**爬虫的网站是<https://www.jd.com/>

需要用到一下库：

import time

import pymysql

from selenium import webdriver

from selenium.webdriver.support import expected\_conditions as EC

from selenium.webdriver.common.by import By

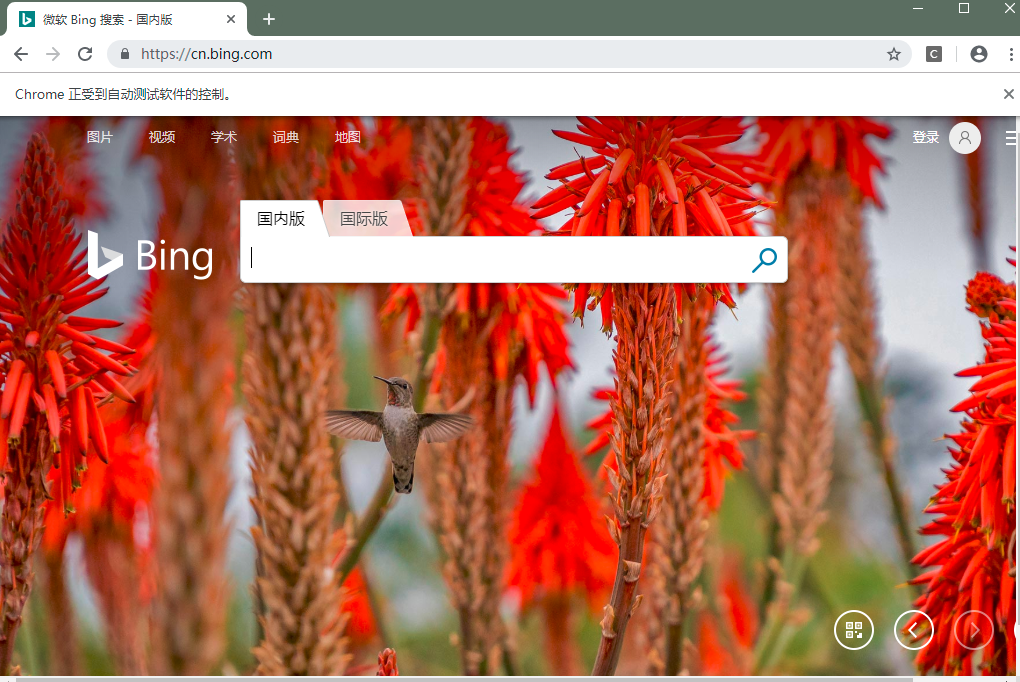
from selenium.webdriver.support.ui import WebDriverWait

from lxml import etree

该动态爬虫主要通过输入相应的关键字，会自动模拟浏览器用户进行相应的点击事件，并下载相关信息。

主要过程如下：

用过chromeDriver驱动模拟浏览器用户打开一个新的浏览器窗口：



然后单击主页中的：京东货物列表



然后自动在搜索框输入关键词：Python

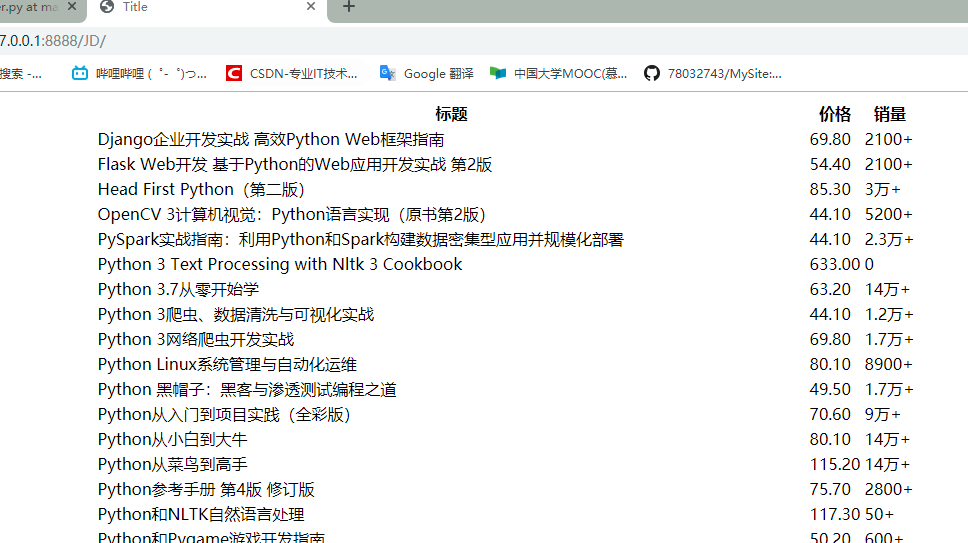


然后模拟用户往下浏览，点击第二页：



由于数据太多电脑不行只模仿的前两页

最后在自己的网页中展示出来：



通过pychram自动创建的相对应的mysql数据库：

