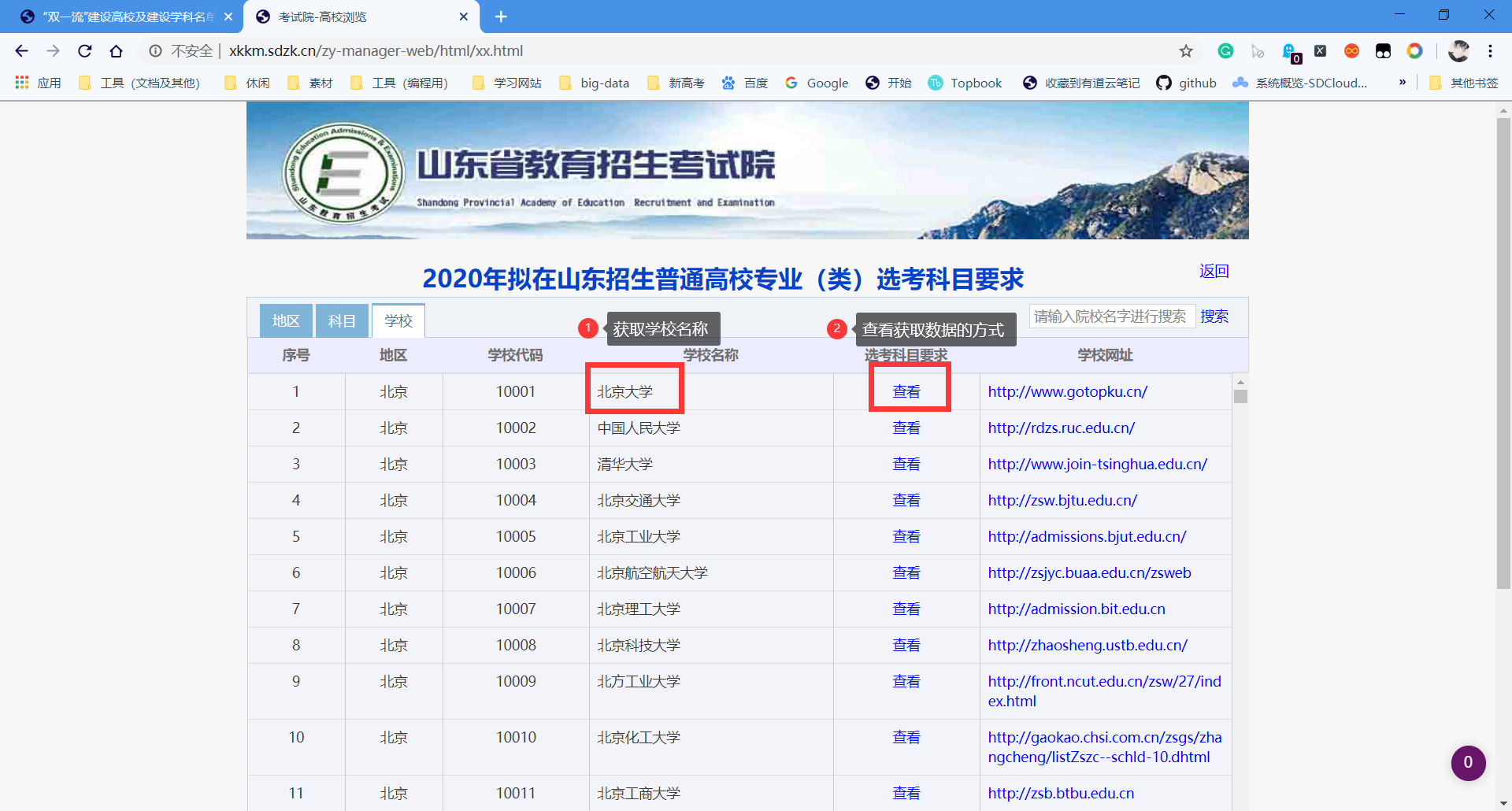
## 

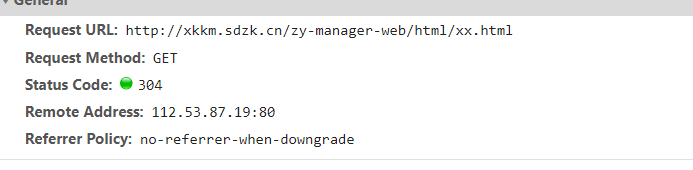
import requests  
from lxml import etree  
import xlwt  
  
#创建Excel文档  
xls = xlwt.Workbook()  
sheet = xls.add\_sheet('选科要求')  
sheet.write(0, 0, '学校')  
sheet.write(0, 1, '专业')  
sheet.write(0, 2, '选科要求')  
# 定义变量k，存放Excel文件中的行  
k = 1  
  
# 各个学校的dm,mc信息网址  
url1 = "http://xkkm.sdzk.cn/zy-manager-web/html/xx.html"  
# 各个学校选科要求的网址  
url2 = "http://xkkm.sdzk.cn/zy-manager-web/gxxx/searchInfor"  
#伪装成浏览器  
head = {  
 'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/79.0.3945.130 Safari/537.36',  
 'Content-Type': 'text/html;charset=UTF-8'  
}  
response = requests.get(url=url1, headers=head)  
response.encoding = None  
html1 = etree.HTML(response.text)  
  
# 利用xpath获取学校名和各学校的选考科目要求的网址、schools用于存放学校名称、scmc用于存放学校的mc属性  
schools = html1.xpath('//div[@id="div5"]//tr/td[4]/text()')  
dms = html1.xpath('//div[@id="div5"]//tr/td[5]/form/input[1]/@value')  
mcs = html1.xpath('//div[@id="div5"]//tr/td[5]/form/input[2]/@value')  
  
# 对每一个学校访问  
for j in range(len(schools)):  
  
 data = {  
 'dm': dms[j],  
 'mc': mcs[j]  
 }  
 headers = {  
 'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/80.0.3987.116 Safari/537.36'  
 }  
 response = requests.post(url=url2, data=data, headers=headers)  
 html = etree.HTML(response.text)  
  
 # xpath匹配专业和选科要求、pors用于存放各个学校的专业、limits用于存放各个专业的选科要求  
 pros = html.xpath('//div[@id="ccc"]//tr/td[3]/text()')  
 limits = html.xpath('//div[@id="ccc"]//tr/td[4]/text()')  
  
 # 输出学校名称  
 print('--------------------' + schools[j] + '-----------------')  
  
 # 写数据  
 for i in range(len(pros)):  
 # 数据清洗  
 pros[i] = pros[i].replace('\r\n', '')  
 pros[i] = pros[i].replace('\t', '')  
 pros[i] = pros[i].replace(' ', '')  
 limits[i] = limits[i].replace('\r\n', '')  
 limits[i] = limits[i].replace('\t', '')  
 limits[i] = limits[i].replace(' ', '')  
  
 # 打印专业、和专业限制  
 print(pros[i], limits[i])  
  
 sheet.write(k, 0, schools[j])  
 sheet.write(k, 1, pros[i])  
 sheet.write(k, 2, limits[i])  
 k += 1  
# 保存数据到Excel中  
xls.save('选科要求.xls')

## 1.选科要求爬虫

1.网站：http://xkkm.sdzk.cn/zy-manager-web/html/xx.html



2.封装headers ，利用requests中的get方法获取学校名



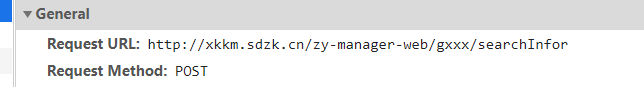
3.获取查看中的dm、mc属性



dms = html1.xpath('//div[@id="div5"]//tr/td[5]/form/input[1]/@value')  
mcs = html1.xpath('//div[@id="div5"]//tr/td[5]/form/input[2]/@value')

4.选科要求的网址、获取其url

# 各个学校选科要求的网址  
url2 = "http://xkkm.sdzk.cn/zy-manager-web/gxxx/searchInfor"



5.创建headers和data

data = {  
 'dm': dms[j],  
 'mc': mcs[j]  
 }  
 headers = {  
 'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/80.0.3987.116 Safari/537.36'  
 }  
 response = requests.post(url=url2, data=data, headers=headers)

5.利用xpath匹配要拿到的数据



pros = html.xpath('//div[@id="ccc"]//tr/td[3]/text()')  
limits = html.xpath('//div[@id="ccc"]//tr/td[4]/text()')

6.打印如下

