**基于python多进程实现POST请求方式网站（拉勾网）的数据爬取及可视化**

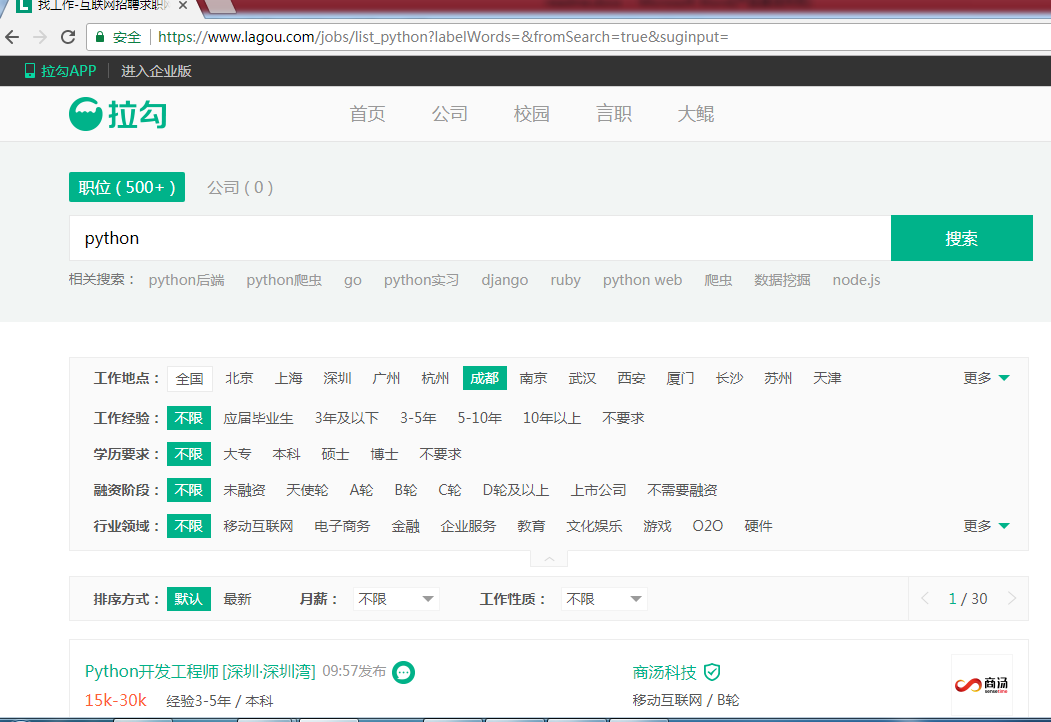
**工具：**

**Python3、pycharm、google浏览器**

**需要安装的库：**

import requests *#请求网站*import pandas as pd *#处理爬取下来的数据格式*import pymysql *#连接mysql数据库*import uuid *#随机生成cookies的值，用于解决反爬*from multiprocessing import Pool *#进程池*from sqlalchemy import create\_engine *#连接数据库*import matplotlib.pyplot as plt *#绘图*import matplotlib as mob *#配置字体*import jieba.analyse *#分词*from wordcloud import WordCloud,ImageColorGenerator *#制作词云图*from PIL import Image *# 打开图片*import numpy as np *# 处理图片*import pylab as pl *#绘制横纵坐标图*

1. **分析网站**
2. **使用谷歌浏览器访问拉勾网主页，随便搜索一个职位名称。**



1. **打开开发者工具，选择Network进行抓包，通过抓包可以知道客户端与服务器之间进行怎样的一种交互。**



**3、刷新网页进行抓包，由于拉勾网的网络请求方式是POST，就是通过json格式的表单提交给服务器，服务器再返回对应的值。所以点击XHR按钮，可以看到提交给服务器的包，点击第一个包，可以看到request url是请求服务器地址，request method 是请求方式。**

