

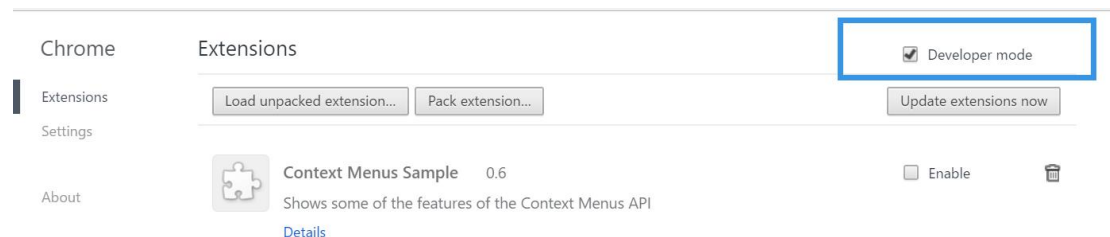
Web Crawler 安装及操作说明

Linkedin 可抓取 post 的 likes 和 comments; 可抓取 1000 条 jobs/company 的 search 数据. 结果均已 csv 格式导出.

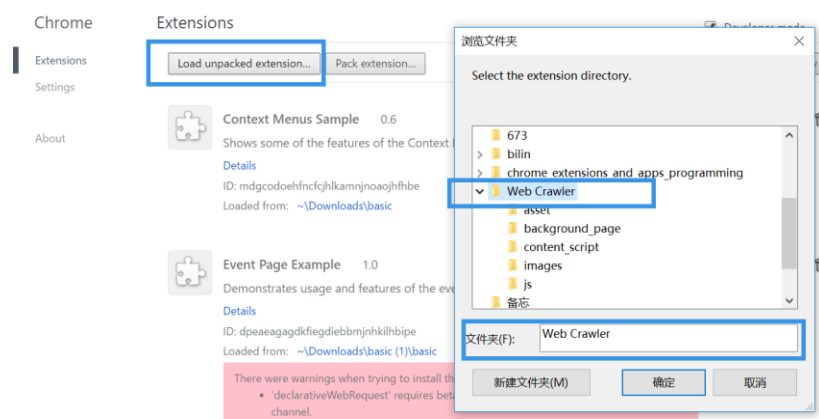
Twitter 可针对一条 tweet, 抓取其 content, likes, comments, retweets. 结果以 json 格式导出.

1 安装

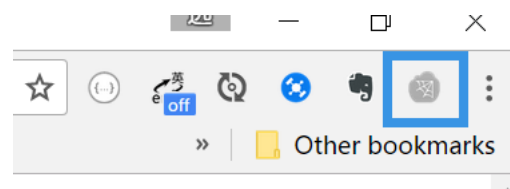
1.1 打开 chrome 设置中的扩展程序 Extensions, 勾选上方的开发者模式 Developer mode;



1.2 点击'加载已解压的扩展程序'/'Load unpacked extension...'; 选择对应的文件夹;



1.3 安装成功后, 浏览器上方会显现对应的图标;

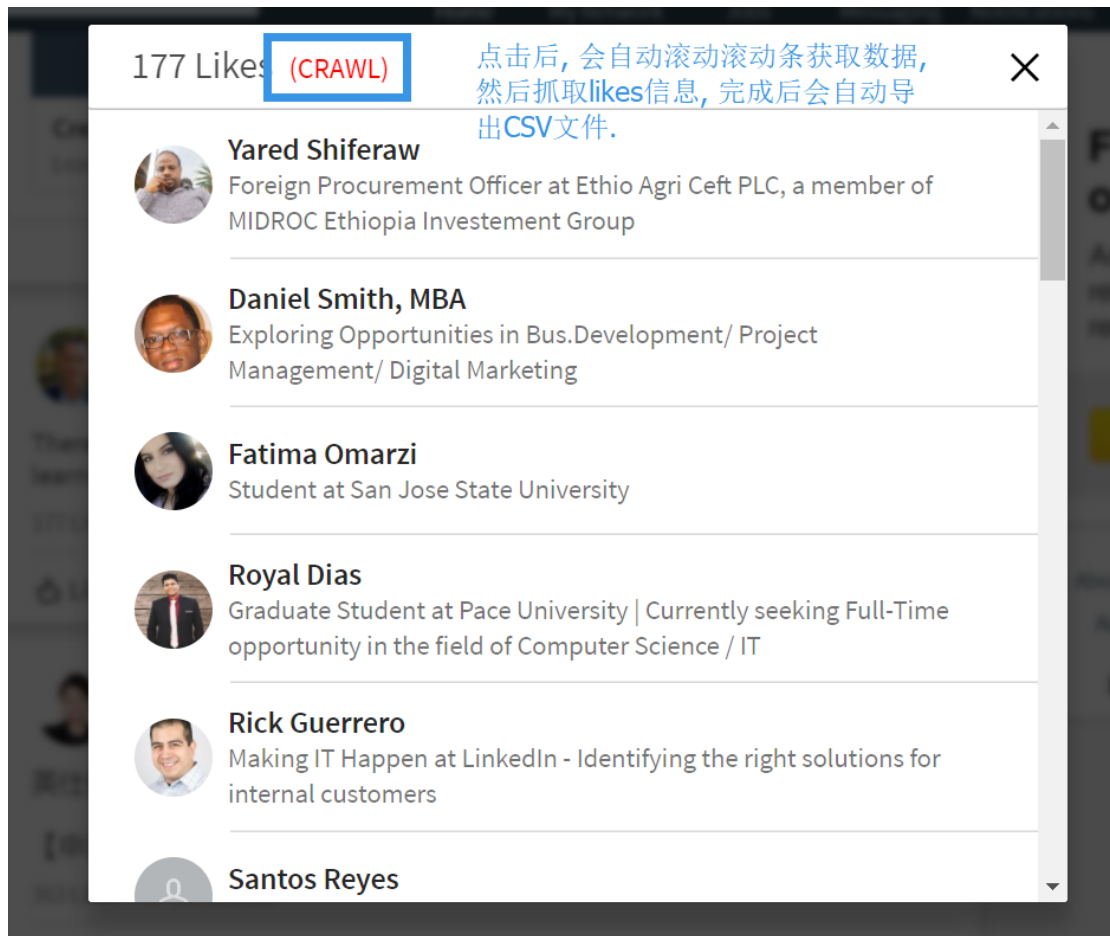


2 操作说明

2.1 Linkedin Post 抓取方法有两种, 推荐使用方法二.

方法一(页面抓取, 抓取 likes)

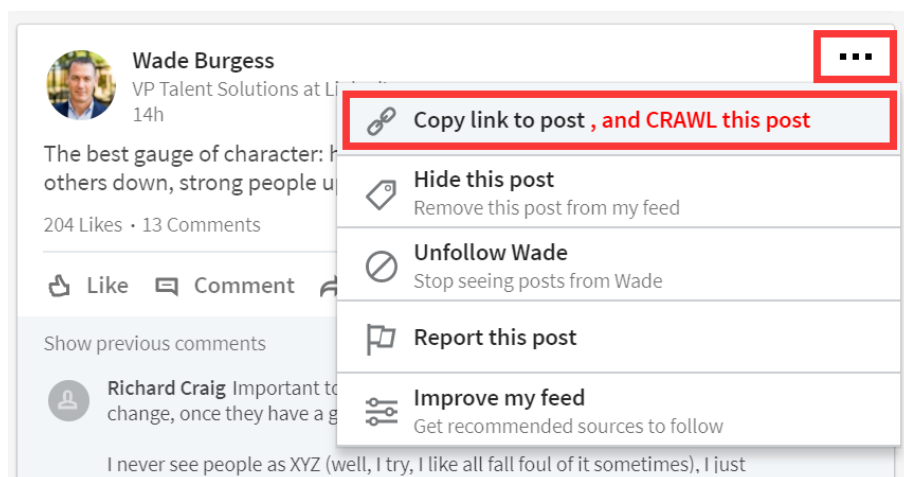
打开 LinkedIn, 点击 likes 打开模态框. 此时 header 部分会显示 crawl 按钮(若无, 请刷新页面或重启浏览器), 点击后将自动抓取所有的 likes 信息. 完成抓取后, 浏览器将自动下载结果.



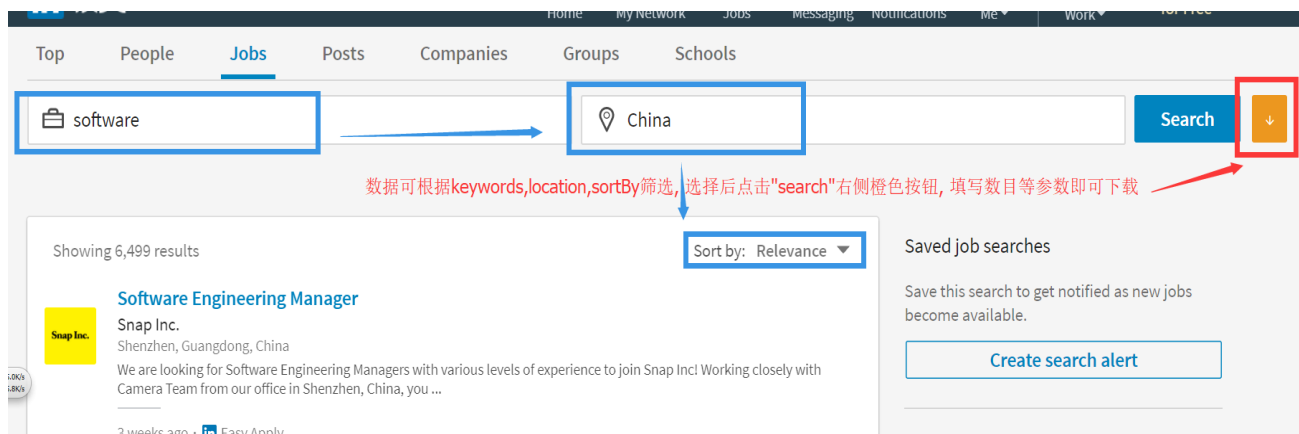
若导出人数与页面显示人数不符, 可手动滚动至页面最底部, 然后再次点击 crawl 按钮.

方法二(API 抓取, 抓取 likes 和 comments)

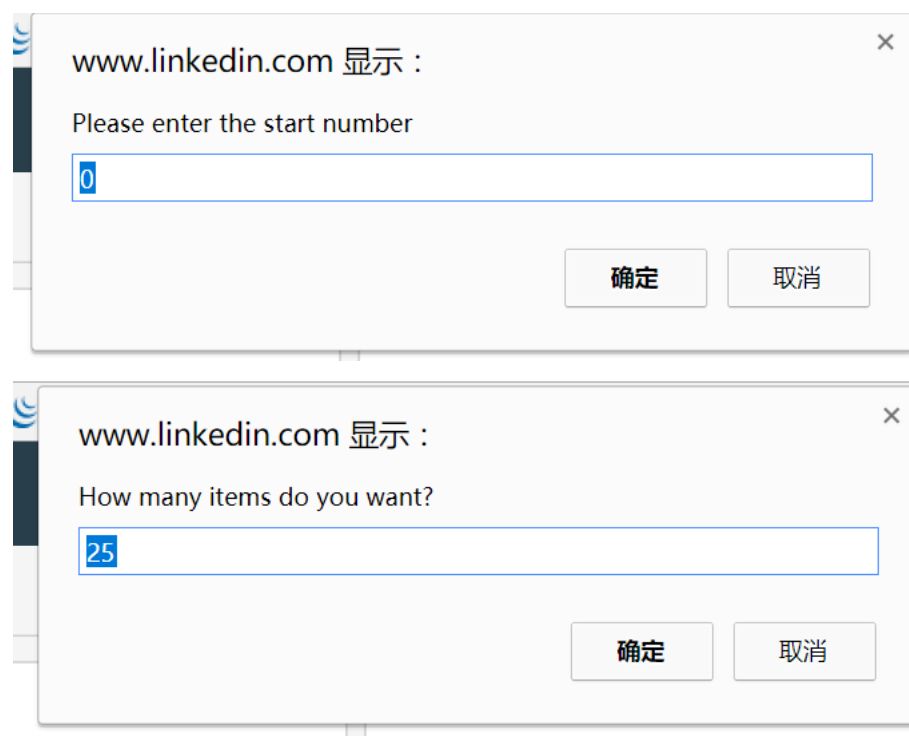
点击 post 右上角按钮, 然后点击“Copy link to post”, 后台将自动下载信息



2.2 LinkedIn Search 抓取(Jobs/Companies)



打开对应的 search 页面, 选择 keywords 等参数, 然后点击橙色按钮
(若无,请刷新页面重试), 填入下载参数即可等待结果导出.



2.3 Twitter 中 tweet 等数据的抓取

打开某条 tweet, 点击页面右侧的橙色按钮, 插件将自动抓取所有信息。数据最终以 json 格式导出。可通过 <http://jsonprettyprint.com/>

查看 pretty 后的数据。

