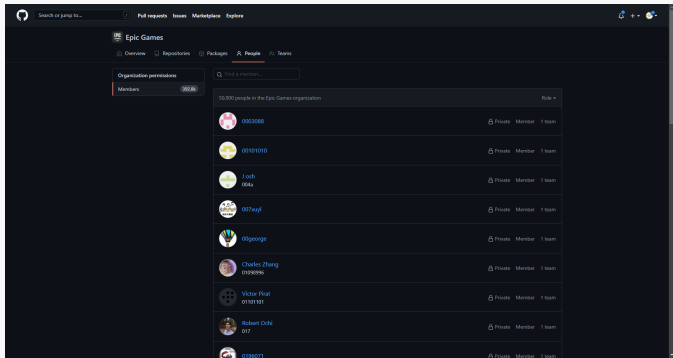


Organization People List Builder

一个半练手性质的 Puppeteer 爬虫，有界面交互逻辑，用来爬取 GitHub Organization 的人员名单与对应所在的页数。目的是想解决由于 Organization 人数过多导致的搜索功能 timeout 报错的问题。

A Puppeteer crawler project with interface interaction logic that crawls the GitHub Organization's people list and their corresponding pages. The purpose is to solve the problem that the search function timeout fails due to too many people in the Organization.

要爬取的 organization 的人员列表		people's list for an organization as the datas this script collect
将人员名与所在的页码相对应		corresponding people's name with the page number
通过这样的格式可以访问指定的页码	<code>https://github.com/orgs/{ORG_NAME}/people?page={PAGE_NUMBER}</code>	view specific page by url in this format
如果直接搜索会得到这样的结果		this is the result for searching it
这是搜索的格式	<code>https://github.com/orgs/{ORG_NAME}/people?query={SEARCH_STRING}</code>	this is the searching format

最终事实证明，超过五万人之后在有限页码内人员本来就只有五万个，超出的不但搜索不到甚至连对应的页数也没有。

In the end, it turned out that there were only 50,000 people in the limited number of pages, and those beyond that could not be searched or even had no corresponding page number.

使用程序 | Run this script

1 运行 `setup.bat` | Run `setup.bat`

`setup.bat` 需要电脑已安装 node, 此 bat 将会使用 npm 安装运行时必要的组件。

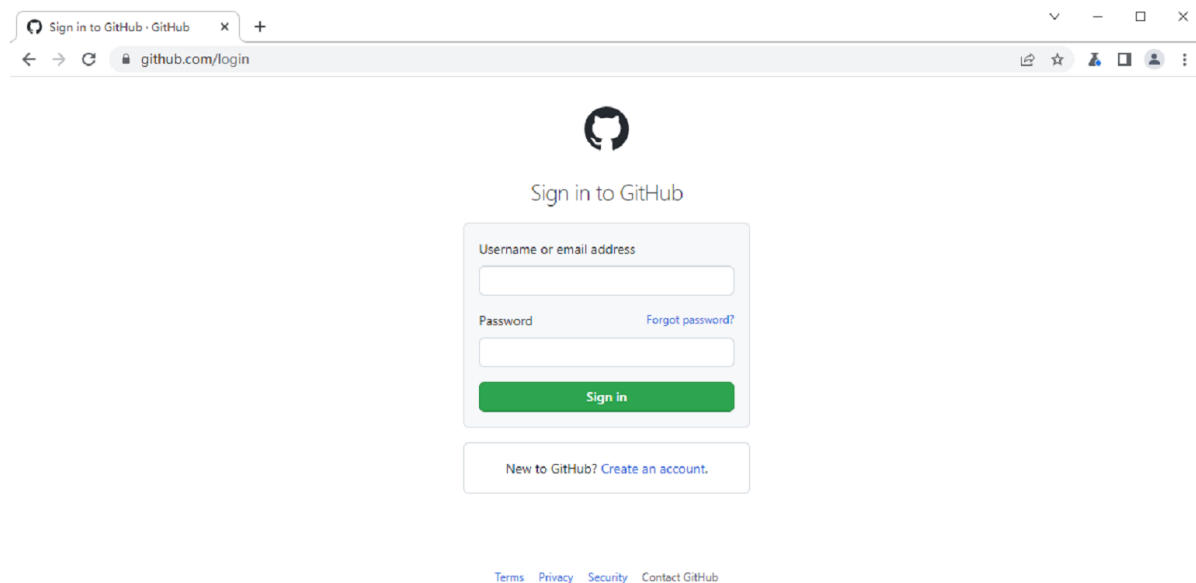
`setup.bat` requires node to be installed on the computer. This bat will use npm to install the necessary components for runtime.

2 运行 `run.bat` | Run `run.bat`

`run.bat` 会用 node 启动程序, 其中一条需要管理员权限以用 tsc 编译并更新 `main.ts` 文件, 如果只是运行的话这里没有更新 typescript 代码所以失败了也不影响。

`run.bat` starts the program by node. One command need the administration to run tsc to update `main.ts`, but while here just need to run it with no update, it's error report by no administration can be ignore.

3 登录 GitHub | Login GitHub



4 操作面板 | Control Panel

Title

文件 | D:/Users/Admin/Documents/GitHub/SynRGB/OrgPeopleListBuilder/index.html

☆

QueryPeopleInOrg

Now give the organization's GitHub ID

eg: If the organization's page link is `https://github.com/SynRGB` , then fill in `SynRGB`

After click the "submit" button, the crawler will begin to collect data automatically.
So you can now minimize the page and wait, or at least don't do anything.

Here will show the databases already formed (if has):

- [EpicGames.csv](#)
- [EpicGames.db](#)
- [HMUniversity.csv](#)
- [HMUniversity.db](#)

运行完会再回到此页面，将列出已生成的数据库和csv表，点击可下载
After the script done, return to this page and shows those formed datas in database and csv format, and can be download by click it

此处输入需要抓取的 organization 的ID名，必须是此账号能访问的 organization
Here is for input the organization's ID, the login account should be able to access this org