

- 1.说明
- 2.环境搭建
  - 2.1conda
    - 2.1.1初始化配置
    - 2.1.2创建环境
    - 2.1.3选择环境
  - 2.2Chrome配置
    - 2.2.1chrome浏览器
    - 2.2.2chromeDriver
  - 2.3Wget配置
    - 2.3.1wget 下载
    - 2.3.2配置环境变量
- 3.文件结构说明
- 4.运行
- 5.最终处理

# 1.说明

---

该代码通过selenium库 获取待下载的pdf链接并存储为txt，后使用wget工具下载对应pdf文件。

白水青风客

bsqfke@gmail.com

## 2.环境搭建

---

### 2.1conda

---

conda 是一个环境管理工具，本人使用 **Anaconda**进行管理

#### 2.1.1初始化配置

anaconda 配置相关流程涉及

1. 下载anaconda 官网<https://www.anaconda.com/download/>
2. 安装anaconda 可以更改保存路径 一路下一步即可
3. 将anaconda添加到环境变量  
[https://blog.csdn.net/weixin\\_43914658/article/details/108785084](https://blog.csdn.net/weixin_43914658/article/details/108785084)
4. 换源（注意新开cmd窗口） <https://zhuanlan.zhihu.com/p/515925073>

## 2.1.2创建环境

---

1. 创建环境，在cmd窗口 `conda create -n scraw python=3.11` 此处的scraw为环境名 可以自己改
2. 激活环境，同上面窗口 `conda activate scraw` or `activate scraw`
3. 安装包，激活后 `conda install -r requirements.txt` 这里的 `requirements.txt` 可以选择将本文给出的 `requirements.txt` 拖入 或跳转到本文件夹目录下打开cmd执行

## 2.1.3选择环境

---

个人使用的是vscode，在py文件编辑界面 右下角即可选择 `interpreter` ,选择为 `scraw` 环境

## 2.2Chrome配置

---

### 2.2.1chrome浏览器

浏览器下载链接 <https://www.google.cn/intl/zh-CN/chrome/>

### 2.2.2chromeDriver

版本 123.0.6312.105 (正式版本) (64 位)

驱动需要与浏览器版本一致,如图所示前三个字段一致即可，下载链接  
<https://getwebdriver.com/chromedriver>

## 2.3Wget配置

---

参考教程: <https://zhuanlan.zhihu.com/p/28826000>

### 2.3.1wget 下载

下载链接 <https://eternallybored.org/misc/wget/>

### 2.3.2配置环境变量

参考教程 <https://zhuanlan.zhihu.com/p/28826000>

## 3.文件结构说明

---

1. scraw.py 爬取url
2. download 根据url调用wget下载pdf
3. main 主函数文件
4. helper 辅助文件
5. totalSet 设置文件

如需要对异常情况进行筛查 请在totalSet中修改checkFlag变量

其余文件说明详细见totalSet.py

## 4.运行

---

环境配置完成后 在main文件运行即可

Symbol.txt 文件需要提前给定 将任务的公司代号覆盖文件即可

其余文件可自动判断生成

## 5.最终处理

---

注意清楚电脑C盘用户目录下的Cache文件夹中的文件 简便办法参考：

<https://zhuanlan.zhihu.com/p/551218511> （不可以删除整个文件夹，只能删除里面的文件）