# 微博爬虫说明

## 整体思路

首先，确定目标用户: 确定几个用户的UID。UID 是微博用户的唯一标识符。

之后，根据UID找到其关注者获取关注者列表: 获取每个目标用户的关注者UID列表。

随后，获取大量的微博UID，从关注者中获取微博UID: 收集所有关注者的UID。

可以使用循环遍历每个目标用户获取的关注者列表，并合并所有关注者的UID。

最后，爬取这些UID发布的动态，获取用户发布的动态: 使用微博用户UID爬虫技术获取每个关注者发布的微博动态。

## 运行步骤

### 前提条件

已安装Python环境

已安装必要的依赖库

拥有有效的微博Cookie信息

### 步骤 1: 在weibo-follow中配置config.json

配置config.json

在项目目录下找到config.json文件。

将你的微博Cookie信息替换到config.json中的对应字段。示例如下：

json

复制代码

{

"cookie": "your\_weibo\_cookie\_here"

}

### 步骤 2: 运行weibo\_follow.py爬取关注者UID

运行weibo\_follow.py

打开终端或命令提示符，导航到项目目录。

运行以下命令：

bash

复制代码

python weibo\_follow.py

该脚本将使用配置文件中的Cookie信息登录微博，并开始爬取指定用户的关注者UID。

爬取到的UID将被保存到user\_id\_list.txt文件中，每行一个UID。

### 步骤 3: 运行webPosts.py读取并爬取UID发布的动态

读取user\_id\_list.txt中的UID

确保user\_id\_list.txt文件中包含已爬取的关注者UID。

每行一个UID，格式如下：

|  |
| --- |
| 1234567890  0987654321 |

运行webPosts.py

打开终端或命令提示符，导航到项目目录。

运行以下命令：

|  |
| --- |
| python webPosts.py |

该脚本将读取user\_id\_list.txt中的UID，并使用爬虫技术获取这些UID发布的动态并整理成需要的用户发帖数据。

## 输出文件

输出文件格式在UserPosts 文件夹中格式：userPostsInformation0-{UID}.json，例子

|  |
| --- |
| userPostsInformation0-2014254514.json |