领英开发文档

杨杰

技术测试

1. 单个账号采集数测试：（测试时间：2017.4.21）

使用同一个账号在一天内访问3000个个人档案没有出现任何限制。（完成）

1. 不用登陆采集：

先用一个账号登录，在去访问百度，接着再去访问这个账号就可以了。

一：完整信息采集方案

账号策略：

账号存储在Linkedin的LinkedinAccount集合里，目前拥有6个字段，password、lastUsedTime、email、userState、isCheckEmail、lastUsedTimeStr。

Password:领英账号密码;

Email:领英账号；

userState：用户目前的状态，两个值except和normal。定时检查except的账号；

LastUsedTime：最后使用这个账号的事件（UTC时间）;

LastUsedTimeStr:字符串时间，便于人读；

isCheckEmail：是否需要人工登陆邮箱检查，如果True，则这个账号在采集的时候不能使用。

测试阶段：

1. 持续时间：

2017-4-21——2017-4-28

1. 过程：

同时登陆3个账号，使用这三个账号无限制的采集数据，所有数据重新开始，将个人链接放入urlNew集合中，将用户信息放入usersNew中。这两个集合都使用person\_website字段作为唯一索引，进行统一（以前使用的是url和account\_id,但是感觉account\_id并不能快速的定位个人信息，而且现在account\_id也很模糊）。Person\_website格式为‘https://www.linkedin.com/in/\*\*\*\*\*\*\*/’。

1. 目的：
2. 验证用户账号是否可以无限制采集；
3. 统一url和users信息，使这两个集合可以很好地统一起来，方便一一对应，以前是没有一一对应的。
4. 验证用户账号可以采集所有个人信息。