TwitterUser采集函数：

1. user\_info:

功能：使用twitter的api，采集用户信息。

输入参数：

host：mongodb的host

database：存放用户信息的database名

collection：存放用户信息的collection名

accountIdList：需要采集的用户ID列表

screenNameList：需要采集的用户screenName列表

（其中，accountIdList与screenNameList，两者必须有一个）

输出结果：将采集的用户信息存储在collection中。

1. user\_avatar:

功能：采集用户头像或背景图片

输入参数：

pictureType：采集图片类型

bashPath：图片本地存放的路径目录。

输出结果：在用户信息的数据库中增加一个字段——profile\_storage\_path，并在本地保存图片。

1. user\_friends:

功能：采集用户的朋友列表。

输入参数：

maxListNum：指定用户朋友列表采集上限（最多5000）

updateClName：代采集朋友列表的用户所在的数据库名。

upsertClName：采集到的朋友IDList将要存入的数据库集合名，如果为None表示不存储。

accountIdList：指定采集用户的accountId列表

screenNameList：指定采集用户的screenName列表

输出结果：得到朋友idList。

1. user\_followers:

功能：采集用户的粉丝列表。

输入参数：

maxListNum：指定用户粉丝列表采集上限（最多5000）

updateClName：代采集粉丝列表的用户所在的数据库名。

upsertClName：采集到的粉丝IDList将要存入的数据库集合名，如果为None表示不存储。

accountIdList：指定采集用户的accountId列表

screenNameList：指定采集用户的screenName列表

输出结果：得到粉丝idList。

1. old\_user\_info:

功能：仅仅只是把以前拥有不完整信息的user重新跑一次，用于获取api原始数据

输入：proxy

输出结果：得到api原始数据。

Tweet采集函数：

1. twitter\_filter:

功能：根据关键字采集实时tweet

输入参数：

keywords：给定关键词列表

输出结果：获取相关推文

1. twitter\_realtime:

功能：实时采集指定用户所发推文

输入参数：

tweetHost：推文数据库的host

tweetDatabase：推文数据库名

tweetCollection：推文数据库集合名

userHost：指定用户数据库的host

userDatabase：指定用户的数据库名

userCollection：指定用户的集合名

输出结果：获取相关推文

1. twitter\_history:

功能：采集指定用户的历史推文

输入参数：

tweetHost：推文数据库的host

tweetDatabase：推文数据库名

tweetCollection：推文数据库集合名

userHost：指定用户数据库的host

userDatabase：指定用户的数据库名

userCollection：指定用户的集合名

输出结果：获取相关推文

1. tweet\_download:

功能：采集推文中的图片和视频。

输入参数：

imgPath：图片存储路径

vidPath：视频存储路径

infoType：采集的类型（图片或视频）

输出结果：获得相应的图片或视频

1. get\_classification\_tweet:

功能：采集分类用户数据库中的各个集合所有用户的200条最新推文。

输入参数：

tweetHost：推文数据库的host

tweetDatabase：推文数据库名

userHost：指定用户数据库的host

userDatabase：指定用户的数据库名

输出结果：获取相关推文

1. get\_realtimeline:

功能：对目标集合中的用户进行轮询，采集用户实时推文。

输入参数：

tweetHost：推文数据库的host

tweetDatabase：推文数据库名

tweetCollection：推文数据库集合名

userHost：指定用户数据库的host

userDatabase：指定用户的数据库名

userCollection：指定用户的集合名

输出结果：获取相关推文