

导出并分析微信聊天记录，生成词云。

5 commits

1 branch

0 releases

1 contributor

Branch: master

New pull request

Create new file

Upload files

Find File

Clone or download

contr4l Update README.md

Latest commit feed966 on 31 May 2018

README.md	Update README.md	last year
ReadCSV.py	Analyze and export python file.	last year
analyze.py	Analyze and export python file.	last year
wxsqlcipher.exe	Micromsg.db decipher tool.	last year

README.md

# MicromsgHistory

写在最前:

某天和女朋友聊天，我说，我觉得你和我在一起这么久，性格变了不少，从微信聊天记录应该可以反映出来，于是疑问产生了——微信聊天记录怎样才能在PC端查看呢？

需要准备的工具:

已root的安卓手机，RE文件管理器，wxsqlcipher.exe（用于解密数据库），python 3.5及相关库（jieba，wordcloud）

操作步骤:

- 获得聊天记录数据库文件

在手机上使用RE文件管理器进入目录“/data/data/com.tencent.mm/Micromsg”，可以发现若干和账户关联的形如20位+长度的文件夹，在其中一个可以找到文件EnMicromsg.db，这就是微信储存聊天记录的数据文件，将其拷贝至容易访问的文件，进而传输至PC端。

- 获取数据库解密码

有关解密码的根据，这篇文章有详细介绍：安卓逆向之旅

手机的IMEI码，在拨号界面输入\*#06#，可以查看本机的IMEI码，对于双卡手机会得到2-3个，全部记录下来

微信的uin码，可以通过RE文件管理器访问“/data/data/com.tencent.mm/shared\_prefs/system\_config\_prefs.xml”查看。也可以通过网页抓包的方式——

- 1、登录微信网页版（推荐使用chrome浏览器，因为下一步是针对chrome浏览器的）
- 2、开始chrome浏览器内置的抓包模式，即在新建的标签页地址栏中输入：chrome://net-internals/#events
- 2.3、在微信中进行一些聊天操作，然后在浏览器的层级上搜索（ctrl+f）“uin”，不出意外可以搜到一个9位的uin码，这个就是你微信号对应的uin码

将IMEI和UIN码连接在一起后，使用MD5算法加密，选择32位小写的加密方式截取前7位，即为数据库密码。

3. 导出聊天记录

使用wxsqlcipher.exe打开EnMicromsg.db文件，提示需要密码，输入上述获得的密码，可以查看数据库中文件。我们选择table为“message”的信息，加载需要一定时间，即可以查看到本机本账号储存的所有微信聊天记录，而后我们有两种处理方式，一是用内部的sql语句导出内容，经测试效率较低。二是导出为csv文件，利用python的csv模块来处理，这里我们选择方案二，File---export---csv file

4. 筛选聊天信息，修正格式

我们要使用csv模块中DictReader的功能。代码参考ReadCSV.py

至此，我们已经将文件以比较漂亮的形式写入了txt文档并可以留存。下面进行词频分析及词云绘制。

## 5. 分析及绘图

首先使用 **jieba** 库来进行分词（有的日常用语词库未包含，但可以自行添加进入字典，具体方式可以查看 **jieba** 的文档），并统计词频，生成词频字典。

然后根据底图或者默认的正方形，绘制专属的词云图~代码参考 **analyze.py**。

到这里，我们的工作已经做完了。