Excel数据表格关联性分析系统

1、预览版版本

选中关联分析的文件有限数量或者只能单选

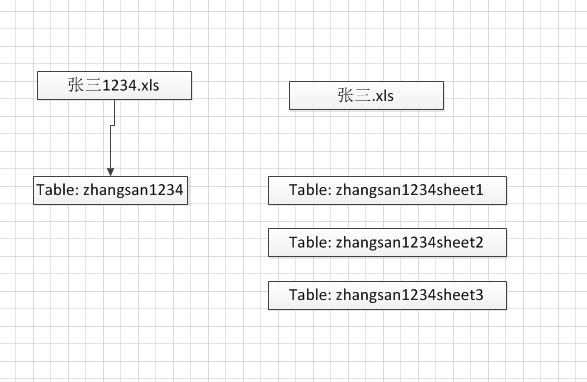
2、

3、

2016/12/24

下一步工作，建立数据库，将excel表格中的数据存储到数据库中，

建立关联表格，



## 一 分析模型

### 1、IP 信息（账户登录日志）

先举个栗子：

QQ账号 A 的登录日志中IP信息： 119.134.12.234 4564 2016-12-26 13:12:54

QQ账号B的登录日志中IP信息：119.134.12.234 4564 2016-12-26 13:13:12

微信账号C的登录日志中IP信息：119.134.12.234 4563 2016-12-26 13:14:54

分析思路：

找到IP地址相同的端口号也要相同，计算时间差是否在运营商的ip回收机制的时间范围内，就可以断定这两个账号是不是在一个移动端登录的

微信账号也同样是这样；

QQ号和微信号之间呢？（这两个应用使用的端口应该是不一样的），但是如果ip是一样的，且都在回收时间内，说明也是在一个移动终端登录的

### 2、好友判断

通过个人信息的好友列表中，是否有对方的账号

### 3、同在一个群

通过个人信息中的 加入的群判断

## 二、Excel数据格式规定

处于算法设计上的考虑，不能达到通用设计，而且系统的重点目标是达到自动分析需要，所以需要给出一个excel数据固定格式

时间格式：

Excel中读取出来的时间数据

1. 如果是excel的时间格式，单元格读取时，需要按照时间读取；
2. 如果是字符串形式的时间，需要按照字符串读取，转换成时间格式

## 三、数据库设计

案件

人员

数据类型（QQ、微信、旺旺信息等）

注意：外键约束 关联的字段长度都需要相同

数据库名称：casedata

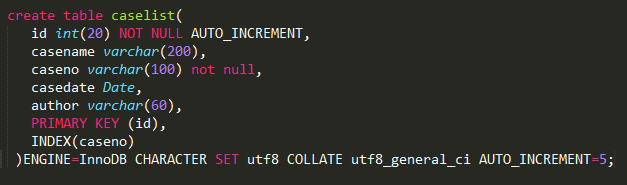
CaseInfo 字段：自增ID，案件名称（含中文名），编号，创建时间，案件创建人

定义如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| id | CaseName | CaseNo | Date | Author |
| 1 | xx小区案件 | MPDQ20170110 | 2017年1月10日 | xxx |
| 2 | xx市区案件 | LSDQ20170110 | 2017年1月10日 | xxx |
|  |  |  |  |  |
|  |  |  |  |  |

例如：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| id | CaseName | CaseNo | Date | Author |
| 1 | 牟平盗窃案 | MPDQ20170110 | 2017年1月10日 | 张队长 |
| 2 | 莱山盗窃案 | LSDQ20170110 | 2017年1月10日 | 吕队长 |
|  |  |  |  |  |
|  |  |  |  |  |



ExcelList 字段：自增id、excel文件名（每个excel文件名）、文件名拼音

|  |  |  |  |
| --- | --- | --- | --- |
| id | FileName | Casetables.CaseNo (foreign key) |  |
| 1 | 13912345423张三 | MPDQ20170110 |  |
| 2 | 15153513456李四 | MPDQ20170110 |  |
| 3 | 15153517894王五 | LSDQ20170110 |  |
| 4 | 15153514962赵四 | LSDQ20170110 |  |

ExcelInfo 表定义如下：

|  |  |  |  |
| --- | --- | --- | --- |
| id | Sheet | ExcelId (foreign key) ExcelList.Id |  |
| 1 | Sheet1 | 1 |  |
| 2 | Sheet2 | 1 |  |
| 3 | Sheet3 | 1 |  |
| 4 | Sheet4 | 1 |  |
| 5 | Sheet1 | 2 |  |
| 6 | Sheet2 | 2 |  |
| 7 | Sheet3 | 2 |  |
| 8 | Sheet4 | 2 |  |
| 9 | Sheet5 | 2 |  |

例子

|  |  |  |  |
| --- | --- | --- | --- |
| id | Sheet | ExcelId (foreign key) PersonsInfo.Id |  |
| 1 | 登陆日志 | 1 |  |
| 2 | 好友列表 | 1 |  |
| 3 | 加入的群 | 1 |  |
| 4 | 注册信息 | 1 |  |
| 5 |  |  |  |
| 6 |  |  |  |
| 7 |  |  |  |
| 8 |  |  |  |
| 9 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| id | Field | SheetId (foreign key) Sheet.id |  |
|  |  |  |  |
|  |  |  |  |

SheetInfo 定义如下：

|  |  |  |  |
| --- | --- | --- | --- |
| id | Field | SheetId (foreign key) Sheet.id |  |
| 1 | IP地址 | 1 |  |
| 2 | IP端口 | 1 |  |
| 3 | 登录时间 | 1 |  |
| 4 | QQ好友 | 2 |  |
| 5 | 微信好友 | 2 |  |
| 6 |  |  |  |
| 7 |  |  |  |
| 8 |  |  |  |
|  |  |  |  |

存储原始数据的表

SheetData 表结构定义

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Id | Field1 | Field2 | Field3 | Field4 | …… |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

表名定义：案件编号 + 文件名 + sheet名称

例如： 表名 MPDQ2017011013912345423ZhangSanDengLuRiZhi

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Id | IPDiZhi | IPDuankou | DengLuShiJian |  | …… |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

存储规则组表（某个案件的所有分析规则） -- 后续扩展

rule\_group

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| id | group | case\_id (foreign key) |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

存储规则清单表

rule\_list

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| id | rule | type | excel\_id (foreign key) |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

例如：

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| id | name | type | excel\_id (foreign key) | case\_no | excel | sheet | field |
| 1 | 寻找13912345423张三QQ好友 | 1 | 1 |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

存储规则明细表

rule\_detial

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id | case\_no | excel\_name | sheet\_name | field\_name | rule\_id (foreign key) |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

例如：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id | case\_no | excel | sheet | field | rule\_id (foreign key) |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

**四、接口设计**

不变化的接口

数据对碰接口 : 输入的数据源和返回对碰结果

数据导入数据库：从excel或别的数据导入mysql或oracle等

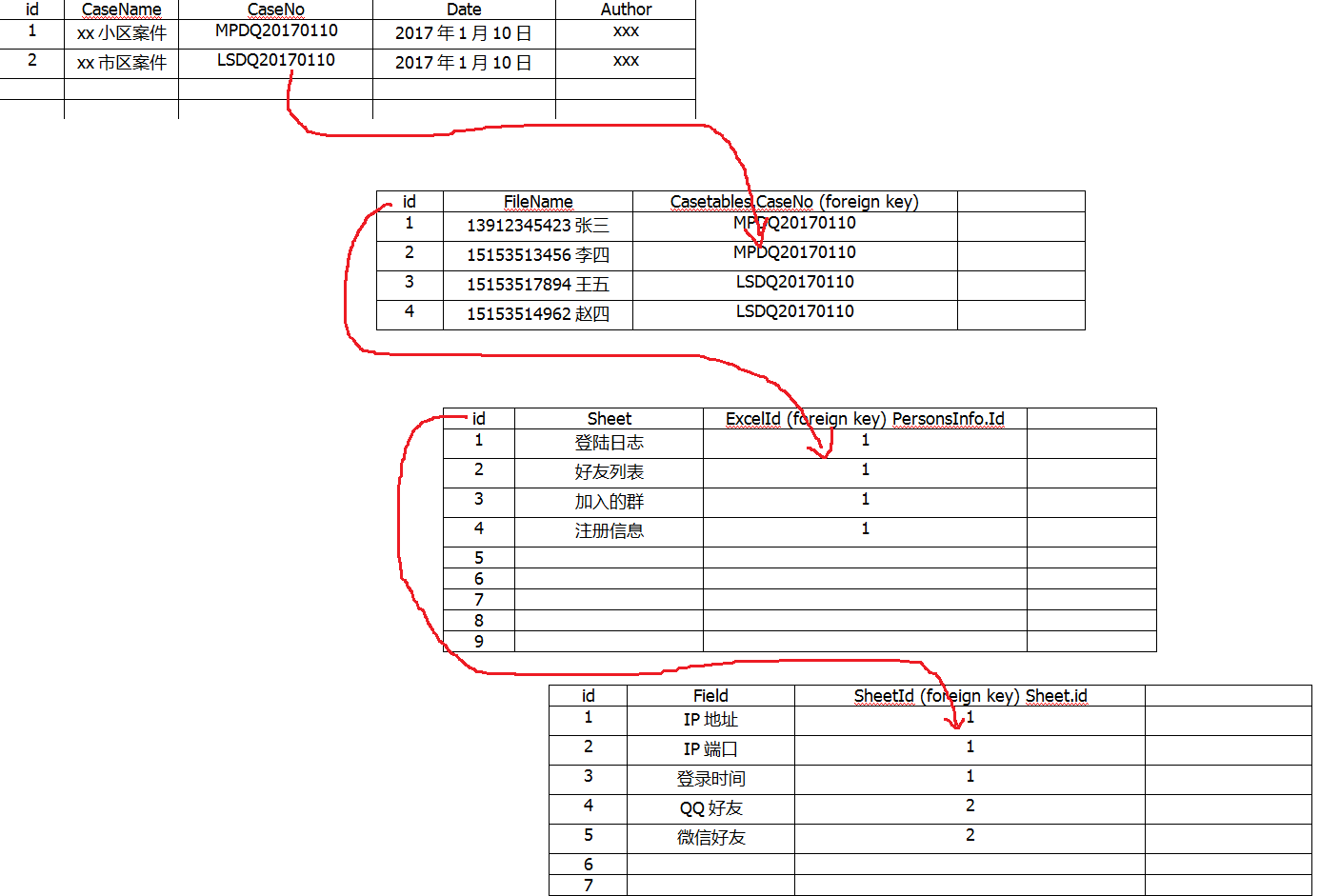
对碰规则入口：

## 五、系统业务流程

1、用户创建“案件” 🡪 导入数据（单个xls文件或压缩包） 🡪

2、导入数据

3、定制数据处理规则（碰撞规则）



根据上图的数据库表格关系，可以知道，提供这个可视化选择过程给用户，用户选择以后可以确定：

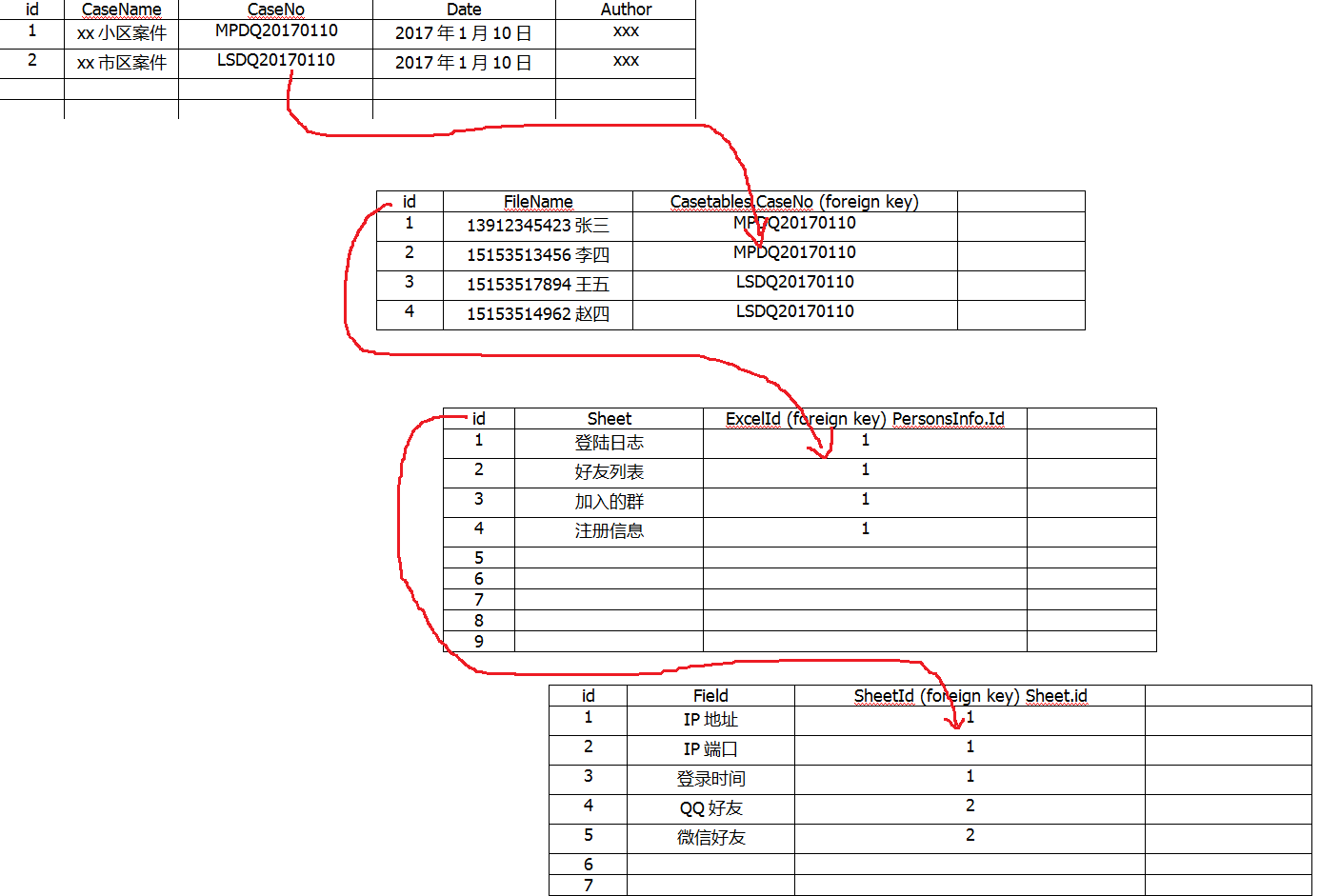
选择待碰撞源； 比如：牟平盗窃案🡪嫌疑人张三🡪注册信息🡪QQ（或者其他如微信）

选择碰撞数据池的对象；（本次要选择的就是和上面已选择好的碰撞数据列）

比如：牟平盗窃案🡪嫌疑人李四🡪好友列表

上面这个过程就是判断 张三的qq是否为李四的好友

4、数据同步



上图显式，当删除某个案件时，相关联的数据都会被删除（由于使用外键约束条件），

也会删除对应的数据表，比如：MPDQ2017011013912345423ZhangSanDengLuRiZhi (目前本功能还未实现)

碰撞规则清单的数据同步

当碰撞清单中