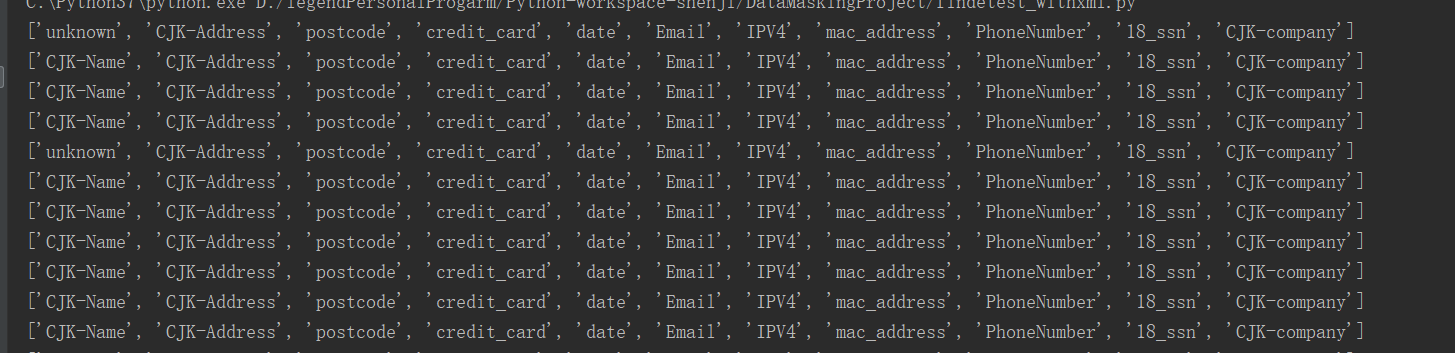
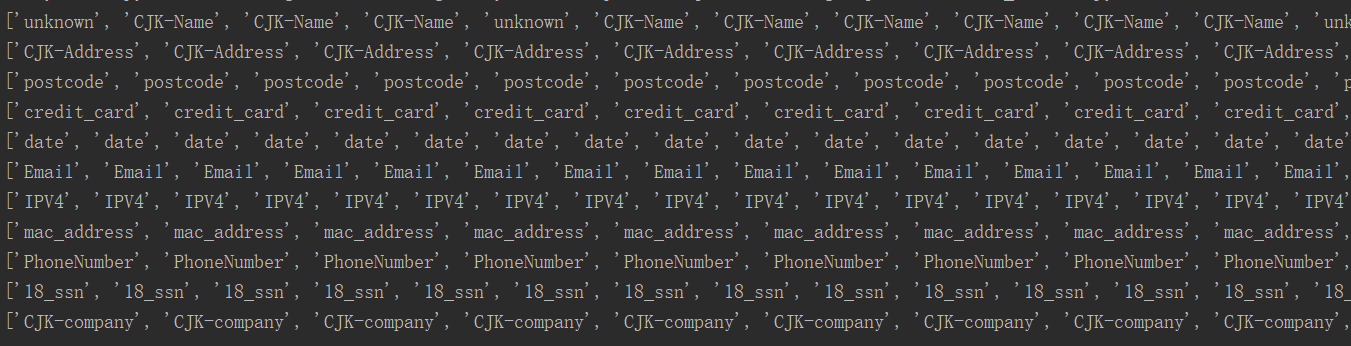
# 准备工作

根据java包，chorine-finder,分为，finder（确定敏感数据源），mask，pattern，

这次主要写的是finder的部分。（pattern就是匹配部分）

Python代码中其实主要就用一个正则去匹配，开始写了一版没有用xml，冗余的代码太多了。所以把java包里的xml直接拿过来了，但是也是直接读文件的方式，去匹配xml的正则，后续看看能不能有其他办法吧

# 源码思路

1. 先fake一些能想得到的敏感数据
2. 
3. 然后识别这些数据的属性，属于姓名，手机号，邮箱等，xml中的正则表达式顺序的原则是越精细的就放到前面，如邮箱，手机号，ip地址，特征越不明显的就放到后面，如qq号，密码等，存到对应的二维列表中
4. 
5. 然后把二维列表的第一列找出来
6. 
7. 去重，作为字典的key，值为零，循环第一列，和key相同就吧值加一。最后得出结果
8. 