k匿名算法

1、显式标识符（ID，能够唯一地确定一条用户记录），如：身份证号、姓名、电话号码等
2、准标识符（QI，能够以较高的概率结合一定的外部信息确定一条用户记录）：单列并不能定位个人，但是多列信息可用来潜在的识别某个人。
3、敏感属性（需要保护的信息）。
4、非敏感属性（一般可以直接发布的信息）。

去识别化是从**数据集中删除显示标识信息的过程**

关联/链接攻击：寻找辅助信息【包含个人标识信息的数据】和去标识数据库的重叠列来重识别个体。

https://blog.csdn.net/qq_41691212/article/details/121739616

“将数据集按照准标识（Quasi-Identifier）分组，使每个分组中的个体都拥有相同的准标识。如果每个分组的大小都至少为 k，则我们称此数据集满足 k-匿名 性”。

“这样一来，虽然攻击者仍然可以将攻击范围缩小至特定 的分组中，但攻击者无法进一步确定分组中的哪个个体才是攻击目标”

验证k-匿名：依次检测数据库的每一行，查看有多少行和当前行的准标识符相同

数据量越少，越难构建立k匿名。

https://blog.csdn.net/qq_41691212/article/details/121742352

n个相互托管的网站，实现一种策略，依次访问得到n组`ip序列`，对于每组`ip`都有至少k组的`ip`权重和相同。

差分隐私

差分隐私（Differential privacy）使用随机应答（Randomized Response）方法确保数据集在输出信息时受单条记录的影响始终低于某个阈值，从而使第三方无法根据输出的变化判断单条记录的更改或增删，被认为是目前基于扰动的隐私保护方法中安全级别最高的方法。

https://blog.csdn.net/S1406793/article/details/127578204