**隐私数据工具集**

**1 隐私数据脱敏处理 测试案例：**

表示需要预先提供测试文件

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **编号** | **数据类型** | **输入** | **结果** | **说明** |
| 1 | 数值 | 数值变量：123.456 | 脱敏后的数值：419.203 | 每一位数字随机替换 |
| 2 | 图片 | **待脱敏文件上传：lena.jpg**  脱敏参数：[230, 240, 115, 130]  高斯核参数：5（默认为5）  输出文件名：lena\_protected.jpg | 脱敏后图片文件lena\_protected.jpg的下载链接（超链接形式） | 图片指定区域  高斯模糊化 |
| 3 | 视频 | **待脱敏文件上传：test.mp4**  脱敏参数：[300, 100, 500, 500]  像素块参数：5（默认为5）  输出文件名：pixelate\_region.mp4 | 脱敏后视频文件pixelate\_region.mp4的下载链接（超链接形式） | 视频指定  区域像素化 |
| 4 | 音频文件 | **待脱敏文件上传：audiotest.wav**  脱敏开始时间：1  脱敏持续时间：1.5  输出文件名：beeped.wav | 脱敏后音频文件beeped.wav的下载链接（超链接形式） | 正弦波替换  音频（哔声） |
| 5 | 实时语音 | 语音录入：我是一名大学生  脱敏标记：大学生（开始-结束标记，选择要脱敏的部分）  输出文件名：audiotest.webm | 脱敏后实时语音文件audiotest.webm的下载  链接（超链接形式） | 正弦波替换  音频（哔声） |
| 6 | ofd | **待脱敏文件上传：ori\_ofd.ofd**  脱敏页码数：[1,2]  输出文件名：protect\_ofd.ofd | 脱敏后ofd文件protect\_ofd.ofd的下载链接（超链接形式） | 指定页面删除 |
| 7 | csv | **待脱敏文件上传：adult\_small.csv**  匿名参数k：3  标识符列号：[0, 1, 2, 3, 4, 5, 6, 7]  敏感列列号：8  输出文件名：adult\_k3.csv | 脱敏后csv文件adult\_k3.csv的下载链接（超链接形式） | k匿名算法实现 |
| 8 | 文字 | 文本输入：hello my name is barry  脱敏词语：barry | 脱敏后的字符串：hello my name is ocawd | 脱敏词语替换为  随机字符串 |
| 9 | 位置 | 目标坐标：(10.3, 10.3)  隐私预算：0.1（默认为0.9） | 脱敏后的坐标：(10.3623, 10.3413) | 添加拉普拉斯噪声 |
| 10 | 轨迹 | 目标轨迹：[(1.0, 2.0), (116.435842, 39.941626), (3.0, 4.0)]  敏感坐标：[(116.435842, 39.941626), (116.353714, 39.939588), (116.435806, 39.908501), (116.356866, 39.907242)]  区域半径：1000（默认为1000）  隐私预算：0.9（默认为0.9） | 脱敏后轨迹坐标：[(0.9234, 2.15132), (117.02431, 39.8412), (3.0312, 4.1245)] | 指定脱敏区域，在区域内时对经纬度坐标添加噪声。 |

**2 隐私保护效果评估 测试案例：**

表示需要预先提供测试文件

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **编号** | **数据类型** | **输入** | **结果** | **说明** |
| 1 | 数值表 | 脱敏前文件：**adult\_small.csv**  脱敏后文件：**adult\_k3.csv** | 信息损失数值 | 信息损失  联合熵做差 |
| 2 | 图片 | 脱敏前文件：**lena.jpg**  脱敏后文件：**lena\_protected.jpg** | SSIM相似度数值 | 结构相似指数SSIM |
| 3 | 视频 | 脱敏前文件：**test.mp4**  脱敏后文件：**pixelate\_region.mp4** | SSIM相似度数值 | 每一帧的SSIM计算平均值 |
| 4 | 音频文件 | 脱敏前文件：**audiotest.wav**  脱敏后文件：**beeped.wav** | 余弦相似度数值 | 余弦相似度 |
| 5 | 实时语音 | 脱敏前文件：**audiofile.webm**  脱敏后文件：**audiotest.webm** | 余弦相似度数值 | 余弦相似度 |
| 6 | pdf | 脱敏前文件：**ori\_pdf.pdf**  脱敏后文件：**protect\_pdf.pdf** | 相似度数值 | 语义相似性 |
| 7 | csv | 脱敏前文件：**ori\_csv.csv**  脱敏后文件：**protect\_csv.csv** | 相似度数值 | 语义相似性 |
| 8 | 文字 | 脱敏前文本：hello my name is john  脱敏后文本：hello my name is allen | 相似度数值 | 语义相似性 |
| 9 | 位置 | 脱敏前坐标：(116.435842, 39.941626)  脱敏后坐标：(1.0, 2.0) | 相似度数值 | 语义相似性  + 数值相似性 |
| 10 | 轨迹 | 脱敏前轨迹：[(1.0, 2.0), (116.435842, 39.941626), (3.0, 4.0)]  脱敏后轨迹：[(1.0, 2.0), (116.43794199999999, 39.932226), (3.0, 4.0)] | 相似度数值 | 语义相似性+ 数值相似性 + 位置关联性 |

**3 隐私数据动态度量 测试案例：**

表示需要预先提供测试文件

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **编号** | **数据类型** | **输入** | **结果** | **说明** |
| 1 | 数值数组 | 数值数组：[3,5,7,9]  敏感数字：7 | 布尔值bool | 是否存在敏感信息 |
| 2 | 图片 | 待测文件：**face.jpg** | 浮点数float | 存在隐私信息的概率 |
| 3 | 视频 | 待测文件：**test.avi** | 浮点数float | 存在隐私信息的概率 |
| 4 | 音频文件 | 待测文件：**test.flac**  敏感词语：America（默认为空） | 浮点数float | 存在隐私信息的概率 |
| 5 | 实时语音 | 实时语音录入：**audiofile.webm**  敏感词语：America（默认为空） | 浮点数float | 存在隐私信息的概率 |
| 6 | pdf | 待测文件：**ori\_pdf.pdf** | 浮点数float | 存在隐私信息的概率 |
| 7 | csv | 待测文件：**ori\_csv.csv** | 浮点数float | 存在隐私信息的概率 |
| 8 | 文字 | 文本输入：I went to the park and saw a dog.  敏感词语：dog | 浮点数float | 敏感词语出现的概率 |
| 9 | 位置 | 待测坐标：(116.435842, 39.941626)  敏感坐标：[(116.435842, 39.941626), (116.353714, 39.939588), (116.435806, 39.908501), (116.356866, 39.907242)]  区域半径：1000（默认为1000） | 布尔值bool | 待测坐标是否存在敏感区域中 |
| 10 | 轨迹 | 待测轨迹：[(1.0, 2.0), (116.435842, 39.941626), (3.0, 4.0)]  敏感坐标：[(116.435842, 39.941626), (116.353714, 39.939588), (116.435806, 39.908501), (116.356866, 39.907242)]  区域半径：1000（默认为1000） | 布尔值bool | 待测轨迹是否存在敏感区域中 |

**需要的文件：**

1. lena.jpg
2. lena\_protected.jpg
3. test.mp4
4. pixelate\_region.mp4
5. audiotest.wav
6. beeped.wav
7. ori\_ofd.ofd
8. ori\_pdf.pdf
9. protect\_pdf.pdf
10. ori\_csv.csv
11. protect\_csv.csv
12. adult\_small.csv
13. adult\_k3.csv
14. face.jpg
15. test.avi
16. test.flac