

目 录

1.引言..... 1

    1.1 编写目的.....1

    1.2 背景.....1

2.软件配置..... 1

    2.1 系统环境需求.....2

    2.2 系统运行.....2

3.用户使用说明..... 3

    3.1 数据库连接模块 .....3

    3.2 水印嵌入模块 .....5

    3.3 水印提取模块 .....9

## 1.引言

### 1.1 编写目的

#### (1) 主要目标

该用户手册的编写目的，是帮助用户在较短时间内迅速掌握 AsterMark 水印软件的使用方法和规范操作。

#### (2) 读者对象

AsterMark 水印软件的测试人员及最终用户。

### 1.2 背景

#### (1) 开发背景

人工智能为大数据的价值挖掘提供了工具，大数据的共享融合更是进一步提升了大数据的价值。然而数据在共享之后由于缺乏有效的追溯和监管手段，容易被共享方再次非法出售以换取二次挖掘的非法利益。因此数据在共享之前，需要针对共享方在数据中增加特定的水印来保障数据的可追溯性，以便于共享方在非法出售数据后，可通过技术手段来鉴别出售数据中的水印，以判别出售者的身份，便于追溯。

#### (2) 研究背景

数字水印技术是 20 世纪 90 年代出现的一门崭新的技术，它通过在数字产品中嵌入可感知或不可感知的信息来确定数字产品的所有权或检验数字内容的原始性，从而在与非法盗用或二次出售者的法律纠纷中获得胜诉。数字水印本质是水印修改了加水印的项，其关键在于尽量最小地改变数据而保持其价值。故针对该系统的数据库数字水印技术应具有以下特点：不可见性、鲁棒性、检测能力。

## 2.软件配置

AsterMark 水印软件采用基于 Java 的 JavaFX 平台进行开发，只需要用户机器上有 Java 运行环境而无需安装即可直接运行。本手册以 Windows 10 为例介绍 AsterMark 水印软件的系统配置。

2.1 系统环境需求

根据项目经验，AsterMark 水印软件最低配置建议如表 2-1 所示。

设备	硬件	需 求
普通 PC 机	CPU	1GHz 处理器或更高
	内存	256MB 或更高
	磁盘	1G 以上

表 2-1 AsterMark 水印软件配置表

2.2 系统运行

在 Windows 环境下，双击软件根目录下的 run.vbs 脚本即可运行。初始界面效果如图 2-2 所示；在 Linux 环境下，在命令行中执行软件根目录下的 start.sh 脚本即可。初始界面效果如图 2-3 所示：



图 2-2

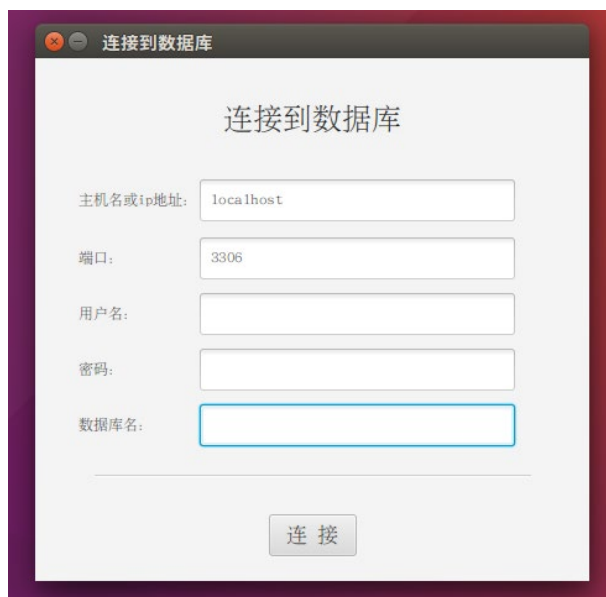


图 2-3

### 3. 用户使用说明

本章详细地讲述了 Astermark 水印软件中最终用户的使用步骤。

#### 3.1 数据库连接模块

(1) **连接数据库**: 运行软件后, 初始连接 Mysql 数据的界面如图 3-1 所示:



图 3-1

其中, 主机名或 ip 地址可以为本地运行的 Mysql 数据库, 即 localhost, 也可以为远程数据库, 即对应的与服务器 ip 地址。

(2) **错误处理：**如果用户没有填写所有必要信息，或者（由于用户名、密码错误等）连接失败，会有如下错误提示：

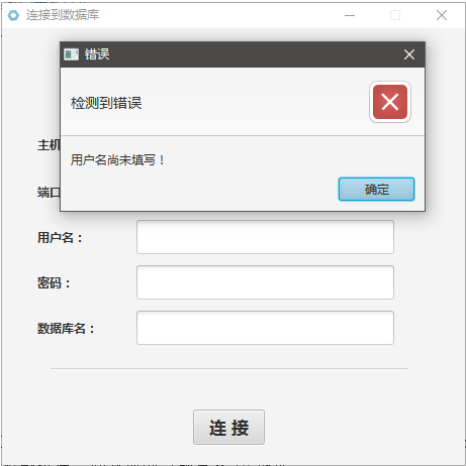


图 3-2

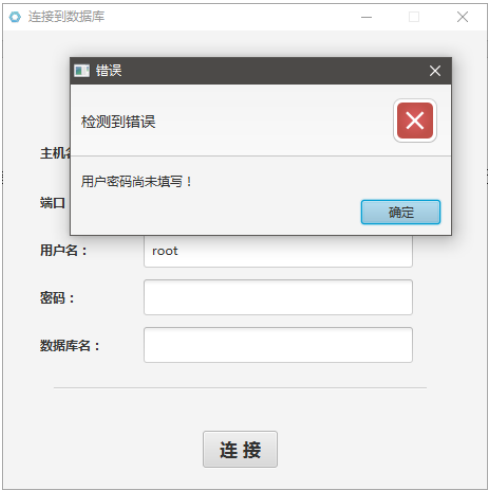


图 3-3

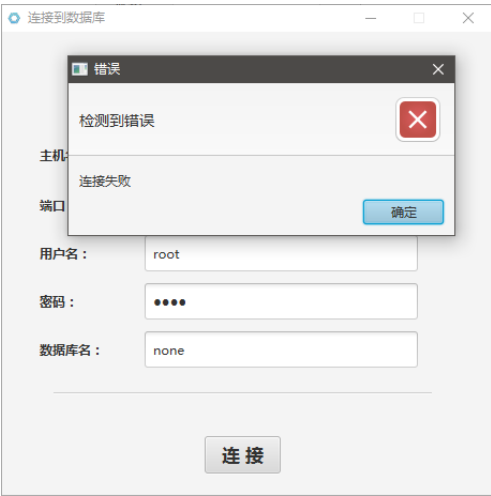


图 3-4

(3) **成功连接：**当连接数据库成功后，若是第一次使用 Astermark 软件连接，会提示将在数据库中自动创建密钥表，用于存储水印的密钥信息，如图 3-5；若已经创建过该表，软件会自动检测到，并直接调至 (4) 功能选择界面：

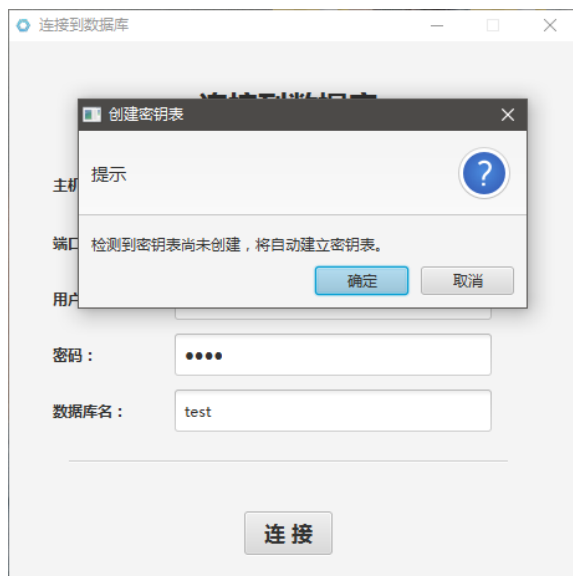


图 3-5

(4) **功能选择界面：**如图 3-6 所示，中间展示的是软件图标以及名称；右上角是帮助按钮，点击将打开该用户手册；下方两个按钮分别对应水印嵌入模块和水印提取模块：



图 3-6

## 3.2 水印嵌入模块

(1) **选择数据表：**在图 3-6 功能选择界面嵌入水印后跳转界面如图 3-7 所示，在左侧列表中会加载出该数据库中存在的的所有表，选中要嵌入的表后，会在右侧显示出该表的具体信息，如图 3-8，方便用户确认：

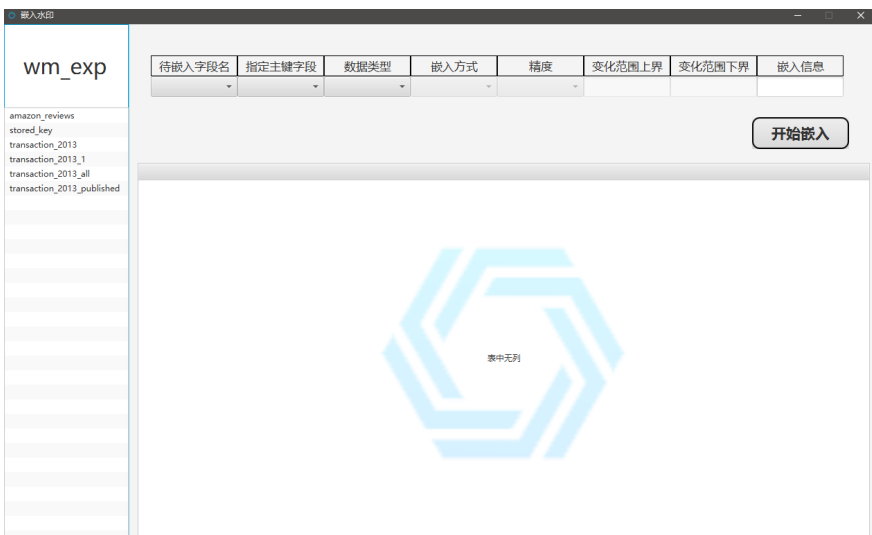


图 3-7

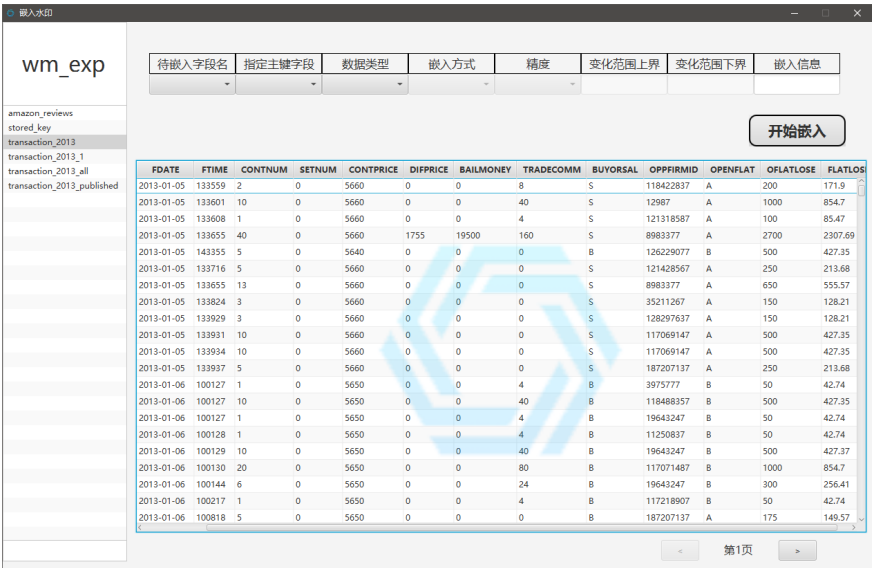


图 3-8

(2) **选择嵌入方式:** 如图 3-9 所示, 这里需要选择待嵌入字段名、指定主键字段, 并选取 (待嵌入字段) 的数据类型——包括整型、浮点型和文本型 (如图 3-10), 其中整型和浮点型 (上两种统称为数值型) 对应的嵌入方式有 LSB 算法和模式搜索算法 (如图 3-11), 文本型对应的嵌入方式有空格嵌入算法、符号嵌入算法和词性逆序数算法 (如图 3-12), 对应具体算法可参考技术文档。

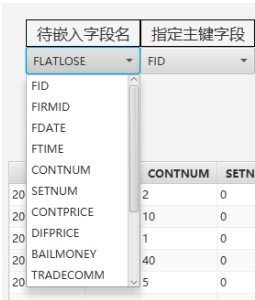


图 3-9

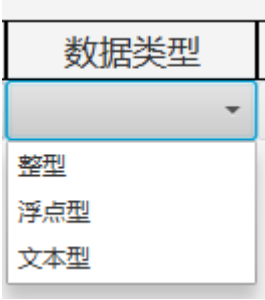


图 3-10

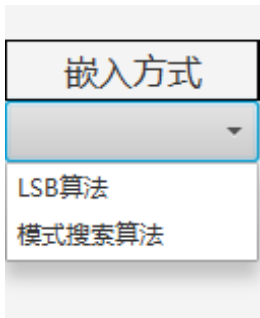


图 3-11

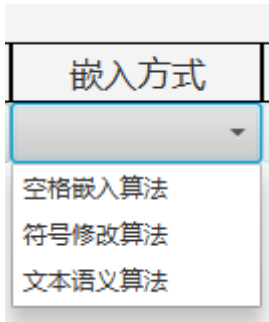


图 3-12

**(3) 输出参数选择：**如果是数值型嵌入，用户可根据实际情况通过精度和变化范围上下界来控制嵌入后数据变化在可接受范围之内，如图 3-13；其中精度为 0.01 代表输出数据精确到小数点后 2 位，0.1 代表小数点后 1 位，1 代表小数点前 1 位，以此类推，如图 3-14 所示：

精度	变化范围上界	变化范围下界
0.01	50	-50

图 3-13



图 3-14

**(4) 嵌入信息填写：**在嵌入信息栏填入数据共享方的信息，如“xx 公司 2019 年 5 月 1 日 数据共享用于用户偏好挖掘”，如图 3-15, 信息的格式、长度不限：

嵌入信息
A公司 2019年5月1E

图 3-15

**(5) 开始嵌入：**点击开始嵌入按钮即开始嵌入水印到数据集中，会有旋转进度条显示，表明当前后台正在进行嵌入，如图 3-16 所示；系统会为该数据表随机生成 16 位的比特串，且保证与该数据表之前因分发给其他对象而已嵌入过的比特串相似度在 0.7 以下（具体计算参考技术文档），方便在溯源时进行区分；最终嵌入完成后嵌入使用的密参信息将存入密钥表 stored\_key 中，如图 3-17：



待嵌入字段名	指定主键字段	数据类型	嵌入方式	精度	变化范围上界	变化范围下界	嵌入信息
FLATLOSE	FID	浮点型	模式搜索算法	0.01	50	-50	A公司 2019年5月1E

开始嵌入

FID	FIRMID	FDATE	FTIME	CONTNUM	SETNUM	CONTPRICE	DIFPRICE	BAILMONEY	TRADECOMM	BUYORSAL	OPPFIRMID	OPENFLAT	OFLAT
1	111111	2013-01-05	133559	2	0	5660	0	0	8	S	118422837	A	200
2	111111	2013-01-05	133601	10	0	5660	0	0	40	S	12987	A	1000
3	111111	2013-01-05	133608	1	0	5660	0	0	4	S	121318587	A	100
4	111111	2013-01-05	133655	40	0	5660	1755	19500	160	S	8983377	A	2700
11	111111	2013-01-05	143355	5	0	5640	0	0	0	B	126229077	B	500
12	111111	2013-01-05	133716	5	0	5660	0	0	0	S	121428567	A	250
13	111111	2013-01-05	133655	13	0	5660	0	0	0	S	8983377	A	650
14	111111	2013-01-05	133824	3	0	5660	0	0	0	S	35211267	A	150
15	111111	2013-01-05	133929	3	0	5660	0	0	0	S	128297637	A	150
16	111111	2013-01-05	133931	10	0	5660	0	0	0	S	117069147	A	500
17	111111	2013-01-05	133934	10	0	5660	0	0	0	S	117069147	A	500
18	111111	2013-01-05	133937	5	0	5660	0	0	0	S	187207137	A	250
19	111111	2013-01-06	100127	1	0	5650	0	0	4	B	3975777	B	50
20	111111	2013-01-06	100127	10	0	5650	0	0	40	B	118488357	B	500
21	111111	2013-01-06	100127	1	0	5650	0	0	4	B	19643247	B	50
22	111111	2013-01-06	100128	1	0	5650	0	0	4	B	11250837	B	50
23	111111	2013-01-06	100129	10	0	5650	0	0	40	B	19643247	B	500
24	111111	2013-01-06	100130	20	0	5650	0	0	80	B	117071487	B	1000
25	111111	2013-01-06	100144	6	0	5650	0	0	24	B	19643247	B	300
26	111111	2013-01-06	100217	1	0	5650	0	0	4	B	117218907	B	50
41	111111	2013-01-06	100818	5	0	5650	0	0	0	B	187207137	A	175

图 3-16

id	watermark	wm_length	threshold	secretKey	secretCode	min_length	db_table	partition_count	target
1	0110010011110110	16	0	0.28	xjzj0	10	wm_exp:amazon_reviews_50000	500	test
2	0000110100011111	16	0.5989505683264753	0.24	0w9uk8	10	wm_exp:transaction_2013	2426	A公司 2019年5月1日 数据共享用于用户偏好挖掘
3	1010000011101001	16	0.5990544395641031	0.24	0w9uk8	10	wm_exp:transaction_2013	2426	B公司 2019年5月1日 test

图 3-17

(7) 嵌入结束：嵌入完成后，会显示嵌入成功后嵌入字段的数据统计变化量（这里以浮点数嵌入为例），如图 3-17 所示，展示了该列数据的均值变化、均值变化比例、方差变化、方差变化比例；并可选择（可多选）嵌入结果的发布渠道——发布到当前数据库、或直接发布为 csv 文件，如图 3-18 所示：

嵌入成功

嵌入完毕!

嵌入后，被嵌入数据的统计量变化如下

均值由 -116.25523 变为 -119.31837  
变化比例为2.63%

方差由 13984438.01140 变为 13983711.75167  
变化比例为.0052%

嵌入结果发布到:

☒ 当前数据库

☐ csv文件

开始发布

图 3-17

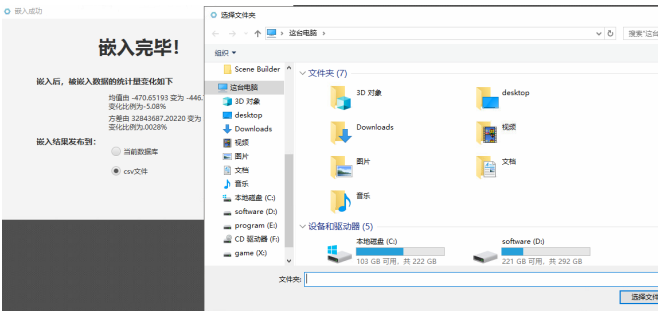


图 3-18

### 3.3 水印提取模块

**(1) 选择数据来源:** 在图 3-6 功能选择界面提取水印后跳转界面如图 3-19 所示。这里有两种导入数据源的方式，一是从数据库导入，二是从 csv 文件导入——若是选择从数据库导入，考虑到这里的数据库可能来自别的主机或别的数据库，而不是当前连接的源数据库，如图 3-20 会弹出新的连接框；若是选择从 csv 文件导入，如图 3-21 会弹出提示框选择 csv 文件头部是否为字段名：



图 3-19



图 3-20



图 3-21

(2) **选择提取参数：**此步的参数均为选择，无需填写。首先系统会查找密钥表 stored\_key 中所有嵌入过的数据库表名（格式为“原始数据库名::原始数据表名”），用户只需选取该表对应的数据库表名，如图 3-22。接着，在选取当前数据表名，（若是从文件导入则只有一个选项），如图 3-23，选取后会在表格中加载出该表的所有数据信息。然后在和之前嵌入步骤相同，选取待提取的字段名（即之前嵌入的字段名）和指定主键字段（与嵌入时指定的主键一致），再选择对应数据类型和提取方式（与嵌入方式一致），如图 3-24。最后点击开始提取，即可：



图 3-22

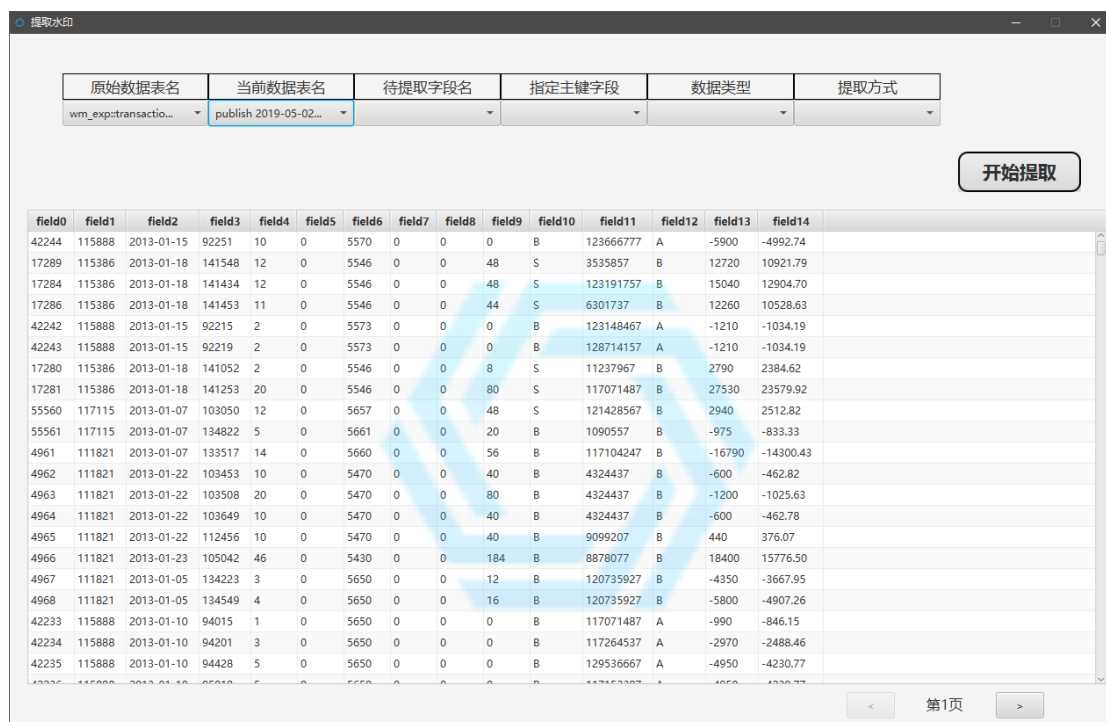


图 3-23

原始数据表名	当前数据表名	待提取字段名	指定主键字段	数据类型	提取方式
wm_exp:transactio...	publish 2019-05-02...	field14	field0	浮点型	模式搜索算法

图 3-24

(3) **提取结果：**提取成功后，会弹出提取结果窗口，如图 3-25 所示，系统会检索密钥表中对应原始数据表所有存在的水印，再根据提取出的水印比特串，去一一进行相似度计算，最后按照相似度从大到小排序将嵌入的目标信息呈现给用户；点击保存按钮可将提取结果保存到文件中，如图 3-26 所示，也可直接点击取消结束提取流程：



图 3-25

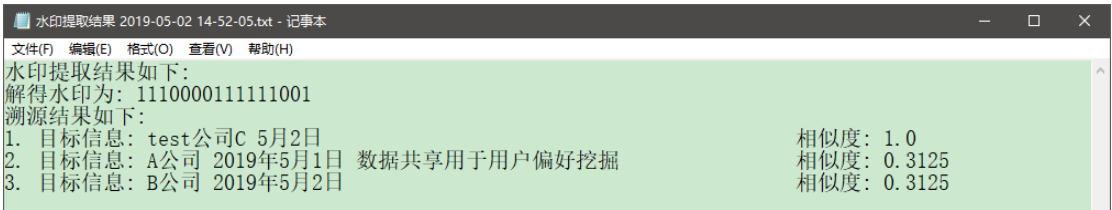


图 3-26