

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ – ПРОЦЕССОВ
УПРАВЛЕНИЯ

ОТЧЕТ
по лабораторной работе №4
по дисциплине «Алгоритмы и структуры данных»
на тему «Деобезличивание данных»
Вариант 13

Выполнила
студентка 2 курса
группы 21-Б15.ПУ
Павлова Ксения Андреевна

Преподаватель
Дик Александр Геннадьевич

Санкт-Петербург
2022

СОДЕРЖАНИЕ

Цель.....	3
Задачи.....	3
Блок-схема общего алгоритма.....	4
Примечания к блок-схеме общего алгоритма.....	5
Рекомендации программиста.....	6
Рекомендации пользователя	6
Контрольный пример	7
Вывод.....	8
Список литературы	9

Цель: написать программу деобезличивания набора данных.

Задачи:

1. Деобезличить первый столбец в таблице – номер телефона;
2. Деобезличить второй и третий столбцы данных в таблице, для каждой строчки которых использован сдвиг буквы по шифру Цезаря, идентичный для одной строки двух столбцов и не обязательно одинаковый для разных строк таблицы;
3. Для каждой строки таблицы вывести шаг, на который были сдвинуты 2 и 3 столбцы.

Блок-схема общего алгоритма.

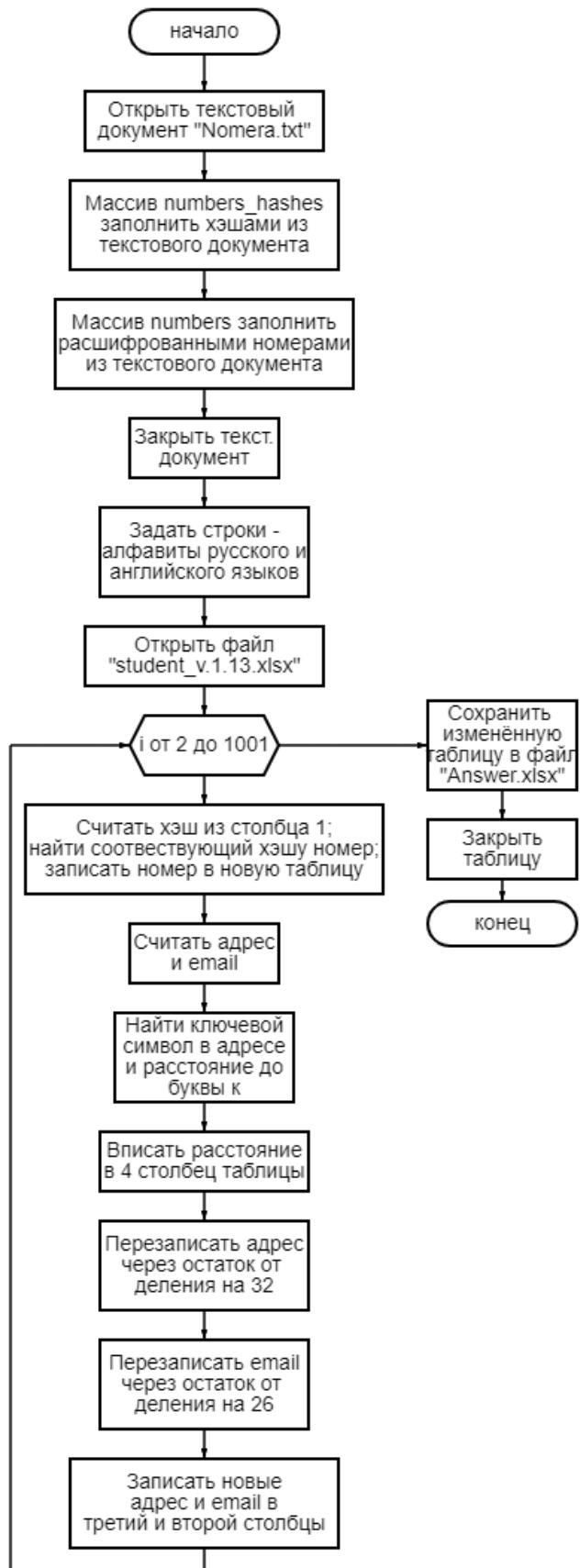


Рис. 1. Блок-схема общего алгоритма

Примечания к блок-схеме общего алгоритма (рис. 1).

Номера телефонов в 1 столбце были зашифрованы с помощью алгоритма хэширования «SHA-1».

Номера телефонов были восстановлены с помощью утилиты HashCat и записаны в текстовый файл «Nomera.txt» в формате «<хэш-значение>:<расшифрованный номер телефона>».

Работа с файлами формата xlsx происходит с помощью библиотеки xlsxwriter.

В алфавите русского языка отсутствует буква ё.

Так как в адресе всегда присутствует номер квартиры и обозначение кв., то ключевой символ – первый символ последней части адреса, разделенного по пробелам. Расстояние (сдвиг) в шифре Цезаря – разность номера ключевого символа и 10 (номер буквы к в алфавите).

Номер символа, который заменяет зашифрованный в адресе, высчитывается как остаток разности номера зашифрованного символа и расстояния по модулю 32.

В случае, когда сдвиг неотрицательный, номер символа, который заменяет зашифрованный в email, высчитывается как остаток разности номера зашифрованного символа и расстояния по модулю 26. Иначе, сдвиг искусственно увеличивается на 32 и номер символа, который заменяет зашифрованный в email, высчитывается по той же формуле, с уже новым расстоянием. Тем самым происходит зацикливание алфавита и нейтрализуется разница в количестве символов в русском и английском алфавите.

В конечный файл вписывается начальный сдвиг.

Рекомендации программиста.

Для работы с кодом необходимо приложение PyCharm, HashCat и установленная библиотека xslxswriter.

Рекомендации пользователя.

Перед запуском программы убедитесь, что в папке с файлом программы находится таблица «student_v.1.13.xlsx» с исходными зашифрованными данными и текстовый документ «Nomera.txt» с хэшами и соответствующими расшифрованными номерами.

Контрольный пример.

В данном разделе представлен пример работы данной программы.

На рис. 2 представлен результат работы программы.

Рис. 2. Вывод результата работы программы

Вывод: в ходе выполнения данной работы была написана программа деобезличивания набора данных.

Список литературы.

1. Introducing Python. Modern computing in simple packages / Bill Lobanovic.
2. <https://habr.com/ru/company/alexhost/blog/536490/>
3. Ссылка на код, необходимые и полученный файлы
<https://github.com/ksenkap/decryption>