Описание способа определения шифра

Значения в столбце «email» сохраняют структуру адреса электронной почты. Можно четко определить, где указаны домены, адреса почтовых серверов, имя пользователя.

Значения в столбце также «Адрес» сохраняют структуру почтовых(городских) адресов. Можно с большой вероятностью предположить, в каких местах указаны сокращения «кв.», «д.».

Так же мы видим, что числа, специальные символы, знаки препинания не перемешиваются.

Из этого выводим предположение: данные зашифрованы каким-то шифром подстановки. Скорее всего таким, где одна зашифрованная буква соответствует одной дешифрованной букве.

Из самых известных этим параметрам соответствуют шифр Цезаря[1]. С него и начнем.

Варианты решения алгоритма

Первый вариант решения – метод частотного анализа[1], когда мы берем статистику для каждого языка, с какой частотой используется та или иная буква и исходя из этого меняем буквы в зашифрованных данных.

Второй вариант – грубый перебор[1].

Третий способ – использовать перебор сдвигов и частотный анализ с известными нам словами(Dictionary Attack[2]) должен работать быстрее чем грубый перебор. И так как он предполагает использование всей известной нам информации о тексте, будем использовать его.

Описание программы

Сет **address_dictionary** -- «известные» нам слова из зашифрованного документа

Функция **caesar_shift_rus** сдвигает буквы русского алфавита на заданное количество шагов.

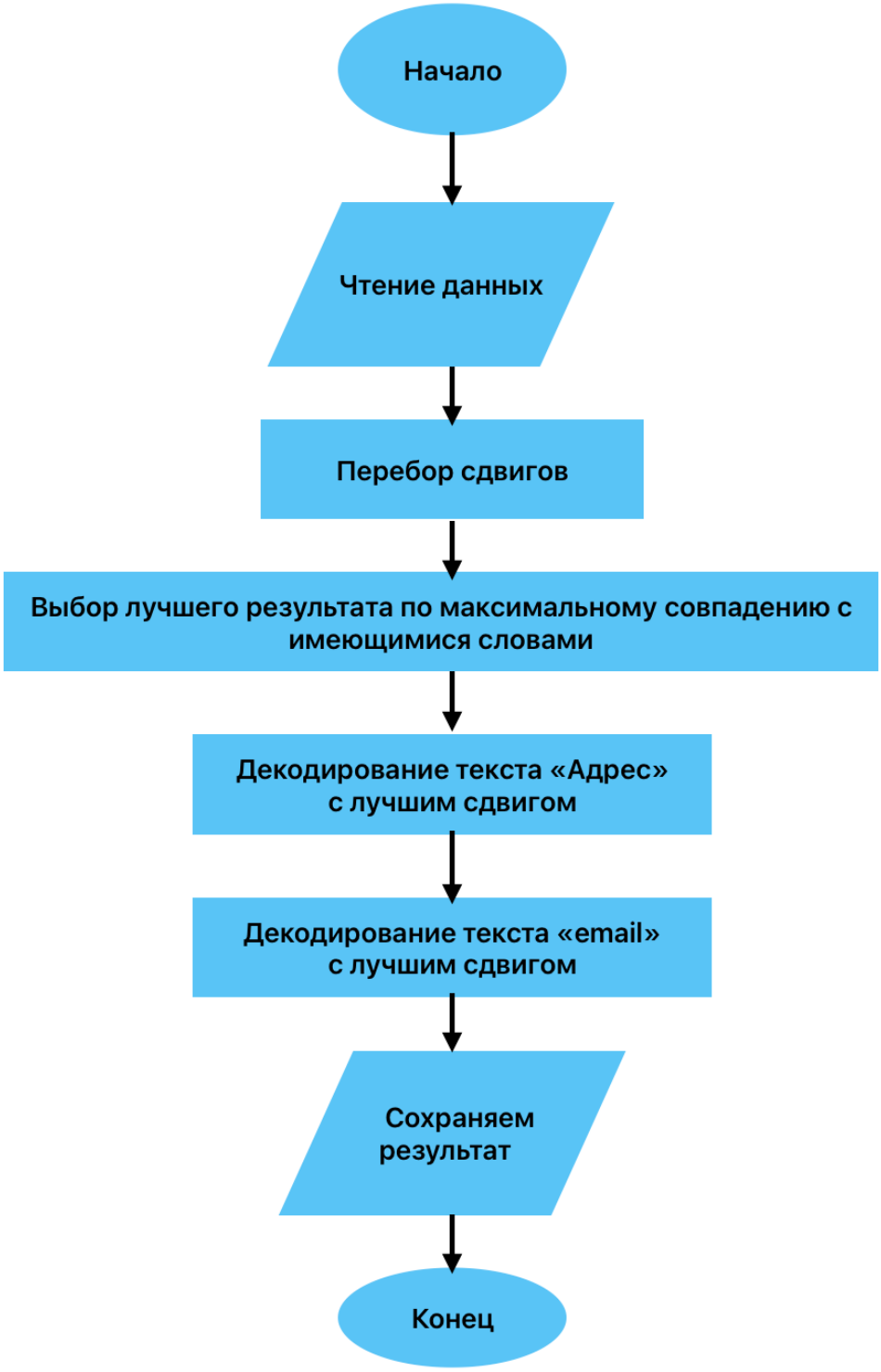
Функция **caesar_shift_eng** выполняет аналогичный сдвиг для английского алфавита.

Функция **score_decoding** проверяет, насколько похож расшифрованный текст на настоящий (по совпадению с заранее «известными» словами из **address_dictionary**).

Функция **decode_line_address** перебирает возможные сдвиги и выбирает тот, который дал наибольшее число совпадений.

Функция **main** читает файл *encrypted.csv* с зашифрованными данными, вызывает выполнение вышеописанных функций и сохраняет результат в файл *decrypted.csv*

Блок схема



Источники

[1] Википедия https://ru.wikipedia.org/wiki/Шифр_Цезаря
[2] Dictionary Attack https://programmador.com/posts/2024/caesar-cipher/#dictionary-attack