Шифр Цезаря

Паначёв Максим Александрович
/ старший преподаватель кафедры ВМиКН /
22 апреля 2020 г.

Кодирование и шифрование

- Кодирование и шифрование информации – близкие по смыслу термины, однако:
  - Кодирование – способ представления информации в каком-либо «удобном» виде;
  - Шифрование – способ защиты/сокрытия информации.
- Закодированный текст может понять каждый, кто знает способ кодирования.
- Смысл зашифрованного текста должен быть ясен только определённым лицам.
- От остальных смысл зашифрованного текста должен быть скрыт и способ его расшифровки без знания секретного кода должен быть как можно более длительным и трудоёмким.

Кодирование и шифрование

- В зависимости от выбранного способа кодирование может как увеличить объём передаваемой информации, так и уменьшить его (а также оставить неизменным).
- Шифрование – в самом лучшем случае оставляет объём таким же, а как правило – увеличивает объём передаваемого сообщения.

Шифр Цезаря

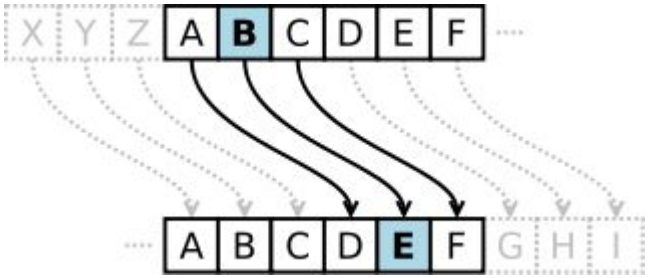
- Рассмотрим простеший способ шифрования, которым человечество пользовалось не одно тысячелетие.
- Шифр (код) Цезаря – шифр простой подстановки.
- Шифр назван в честь римского императора Гая Юлия Цезаря, использовавшего его для секретной переписки со своими генералами.

Шифр Цезаря

- Шифр простой подстановки – шифр, в котором каждой букве исходного текста (plain text) сопоставлен отдельный символ зашифрованного сообщения (encrypted text).
- В общем случае сопоставление осуществляется с помощью таблицы соответствий, которую, очевидно, необходимо держать в тайне.

Шифр Цезаря

- Шифр Цезаря использует **циклическую подстановку**: берём две ленты на которых бесконечно повторяется алфавит нашего сообщения. Шифр простой подстановки определяется путём сдвига одной ленты относительно другой на заданную величину, называемую «числом сдвига».



Шифр Цезаря

- Величина сдвига намного короче таблицы соответствий, её легко запомнить.
- В современных алгоритмах шифрования шифр Цезаря используют как один из шагов сложной многоступенчатой схемы преобразований. При этом шифр может применяться на нескольких этапах несколько раз с различным числом сдвигов

Взлом шифра Цезаря

- Для взлома шифра Цезаря используется простой механизм из теории частотного анализа (раздел криптографии).

Взлом шифра Цезаря

- Шаг 1. Строим таблицу частот символов зашифрованного сообщения **LT**. Таблица частот – это список соответствий – отображение $\Omega \to [0; 1]$, где Ω – Алфавит сообщения (упорядоченный набор уникальных символов):

$$\text{символ} \to \frac{\text{число повторений символа в тексте}}{\text{общее число символов в тексте}}$$

Взлом шифра Цезаря

- Шаг 2. Берём глобальную таблицу частот символов **GT** (например, таблицу частот русского языка) и пытаемся определить величину сдвига исходя из решения следующей задачи:

$$\text{СДВИГ}^* = \underset{0<\text{СДВИГ}<|\Omega|}{\arg\min} \sum_{i=0}^{|\Omega|} \left(GT\left[s_{i+\text{СДВИГ}}\right] - LT[s_i]\right)^2$$

$|\Omega|$ – мощность Алфавита (число символов в Алфавите).

Контрольные вопросы

1. Выберите верное определение понятию «шифр простой подстановки»:
   - Шифр простой подстановки – шифр, в котором каждой букве исходного текста сопоставлен отдельный символ зашифрованного сообщения.
   - Шифр простой подстановки – шифр, в котором каждой букве исходного текста сопоставлена другая буква этого же текста.
2. Закодируйте следующий текст с помощью шифра Цезаря со сдвигом 3: мама мыла раму.
   - йэйэ йшиэ нэйр
   - пгпг пющг угпц
   - умар алым амам
   - овов оэнв твах

Постановка задачи

1. Написать программу кодирующую произвольный текстовый файл в текстовый файл, зашифрованный по алгоритму Цезаря с заданным сдвигом. Обязательные входные параметры программы: имя исходного файла, имя файла с результатом, сдвиг. Необязательные параметры: строка, содержащая символы-исключения, игнорируемые программой при шифровании (т.е. исключаемые из зашифрованного текста).

Обратите внимание: порядок символов в алфавите исходного сообщения устанавливается не по мере появления новых символов в тексте, а в порядке возрастания кодов символов.

Постановка задачи

2. Написать программу формирующую файл, содержащий глобальную таблицу частот (формат хранения определяете самостоятельно, например, JSON). Программа получает на вход путь к каталогу с текстовыми файлами («корпус текстов»). Глобальная таблица частот строится по всей совокупности текстовых файлов в указанном каталоге.

3. Написать программу «угадывающую» сдвиг шифра Цезаря и расшифровывающую исходный файл. Входные параметры: имя зашифрованного файла, имя файла с глобальной таблицей частот, имя файла с результатом. На консоль выводится «угаданный» сдвиг.