

# Блочно-ориентированное сжатие

**Метод: Кодирование длин серий** (*Run-length encoding, RLE*) или **Кодирование повторов**

При сжатии (кодировании) строка одинаковых символов, составляющих серию, заменяется строкой, которая содержит сам повторяющийся символ и количество его повторов.

**Пример 1.** Чёрный текст (черный пиксел – В) на белом фоне (белый пиксел – W)

## Вход:

[illegible]

## Выход:

$$12W1B12W3B24W1B14W \quad (18)$$

# Метод Бэрроуза — Уилера

(Burrows-Wheeler Transform, **BWT**)

- Впервые опубликован в 1994 г. Авторы - Дэвид Уилер и Майк Бэрроуз.
- Сам по себе, BWT не является классическим алгоритмом сжатия, однако выходная последовательность гораздо удобнее для сжатия, нежели исходная.
- Преобразование Б-У применяется в алгоритмах сжатия качественных данных.
- Является компромиссным между быстрыми словарными алгоритмами и статистическими алгоритмами.
- Обычно применяется совместно со статистическими алгоритмами.
- Оперирует целым блоком данных длиной **k** символов

# Сущность BWT

- 1) выделить блок данных из непрерывного потока вх дан
- 2) из полученного блока данных создать матрицу  $W_1$  всех возможных его циклических перестановок; **первой строкой строкой матрицы будет исходная последовательность, второй строкой - она же, сдвинутая на один символ влево, влево, и т. д.**
- 3) отсортировать все строки  $W_1$  в соответствии с лексикографическим порядком символов; получим  $W_2$

**Результат преобразования:** символы последнего столбца  $W_2$  ( $w_k$ ) и номер ( $z$ ) исходной строки среди отсортированных.

Важнейшее свойство алгоритма — **рекуррентность** ( **Рекурсивная функция** (от лат. *recursio* — возвращение) — это числовая функция  $f(n)$  числового аргумента, которая в своей записи содержит себя же. Такая запись позволяет вычислять значения  $f(n)$  на основе значений  $f(n-1)$ ,  $f(n-2)$  ..., подобно рассуждению по индукции.

**Пример 2.** Вход: ЛОГОВО, выход( $w_k$ ): ОООГЛВ

**Пример 3.** Вход: АБРАКАДАБРА, выход ( $w_k$ ):  
РДАКРААААББ

Главной проблемой в реализации BWT является выбор быстрого алгоритма сортировки данных с большой длиной ключа.

Чем меньше мощность алфавита, тем эффективнее метод сжатия по **BWT**

**Пример 3.** Вход: 101010 , выход = ?

BWT используется в архиваторах класса **ZIP** совместно с другими методами