

ЛЕКЦИЯ 5. НЕРАВНОМЕРНОЕ КОДИРОВАНИЕ ДЛЯ ПОСЛЕДОВАТЕЛЬНОСТИ СООБЩЕНИЙ

5.1. Арифметическое кодирование

5.2. Словарные методы сжатия

5.1. Арифметическое кодирование

Алгоритм кодирования:

1. Определить вероятность каждого из символов алфавита в сообщении и назначить каждому из них интервал, пропорциональный его вероятности.
2. Составить таблицу соответствия между символом и интервалом.
3. Установить начало интервала $H_1 = 0$.
4. Установить конец интервала $K_1 = 1$.

5. Выполнять до тех пор, пока есть входные символы
6. Получить входной символ
7. Выбрать из таблицы соответствия начало H_c и конец K_c интервала для символа.
8. Сузить исходный интервал до нового
$$H_i = H_{i-1} + H_c(K_{i-1} - H_{i-1}),$$
$$K_i = H_{i-1} + K_c(K_{i-1} - H_{i-1}).$$

Конец цикла

9. Выбрать конечное значение для передачи
$$X = H_i.$$

Алгоритм декодирования:

1. Получить декодируемое число X_1
2. Выполнять до тех пор, пока декодируемое число X_i не равно «0».
3. Найти символ, для которого начало интервала $H_i \leq X_i$, а конец интервала $K_i > X_i$.
4. Выдать символ.
5. Определить следующее число
$$X_{i+1} = \frac{X_i - H_i}{K_i - H_i}.$$
6. Конец цикла.

5.2. Словарные методы сжатия

Алгоритм LZ78

Просмотренный текст разбирается на фразы, где каждая новая фраза есть самая длинная из уже просмотренных плюс один символ. Она кодируется как номер указателя, где эта фраза встречалась плюс дополнительный символ. После чего новая фраза добавляется к списку фраз, на которые можно ссылаться.

Алгоритм LZW

Кодирование

Последовательно считываются символы входного потока (строка) и проверяется, есть ли в уже созданной таблице такая строка.

Если строка есть, то считывается следующий символ, а если такой строки нет, - заносится в выходной поток код для предыдущей найденной строки, заносится эта строка в таблицу и начинается поиск снова.

Декодирование

LZW-декодер, обрабатывая входной поток закодированных данных, восстанавливает из него исходные данные. Так же, как и алгоритм сжатия, декодер добавляет новые строки в словарь всякий раз, когда находит во входном потоке новый код. Все, что ему остается сделать, — это преобразовать входной код в выходную строку символов и отдать ее на выход декодера.