Практика №3. Построение префиксного кода с помощью алгоритма Хаффмена. Кодирование и декодирование сообщения.

ТЕОРЕТИЧЕСКАЯ ЧАСТЬ.

При построении оптимальных кодов алгоритм Хаффмена не хуже алгоритма Шеннона-Фано.

Алгоритм Хаффмана.

- 1. Последовательность сообщений данной вероятностной схемы располагается в порядке убывания вероятностей.
- 2. Последние два символа объединяются в один с вероятностью, равной сумме вероятностей объединенных символов.
- 3. С полученной последовательностью произвести действия пунктов 1 и 2, до образования последовательности из одного символа с суммарной вероятностью равной 1.
- 4. Строится кодовое дерево, в корне которого стоит символ с вероятностью 1.

ПРИМЕР

Задание. Произвести статистическую обработку данного сообщения, считая, что источник сообщений периодически, достаточно долго выдаёт следующую последовательность символов 12342334551233. Определить энтропию, приходящуюся в среднем на одну букву, длину кода при равномерном кодировании и избыточность. Построить схемы алфавитного кодирования методом Хаффмана. Найти среднюю длину элементарного кода, эффективность сжатия.

Статистическая обработка приведённого сообщения, была выполнена в предыдущем

примере, где и была получена вероятностная схема

X	1	2	3	4	5	\sum_{i}
n	2	3	5	2	2	14
W	2	3	5	2	2	1
	$\overline{14}$	$\overline{14}$	$\frac{\overline{14}}{14}$	$\overline{14}$	$\overline{14}$	

Построим схему кодирования по алгоритму Хаффмена.

		megnipezanini ne ani epini	7 11
символ	P		код
3	$\frac{5}{14}$	/	1
2	$\frac{3}{14}$	5/14 1	011
1	$\frac{2}{14}$		010
4	$\frac{2}{14}$	9/14	001
5	$\frac{2}{14}$	4/14	000

Средняя длина кодового слова равна

$$\bar{l} = \frac{5}{14} + \frac{3}{14} \cdot 3 + \frac{2}{14} \cdot 3 \cdot 3 = 2.29$$

Коэффициент эффективности равен

$$\eta = \frac{2.21}{2.29} = 0.97$$

ЗАДАНИЕ

Провести статистическую обработку текста.

Определить энтропию, приходящуюся в среднем на одну букву, длину кода при равномерном кодировании и избыточность.

Построить схему алфавитного кодирования для однобуквенных сочетаний методом Хаффмана. Найти среднюю длину элементарного кода, эффективность сжатия. Закодировать текст. Декодировать текст.

Построить схему алфавитного кодирования для двухбуквенных сочетаний методом Хаффмана. Найти среднюю длину элементарного кода, эффективность сжатия, сравнить с результатами для однобуквенных сочетаний. Закодировать текст. Декодировать текст.