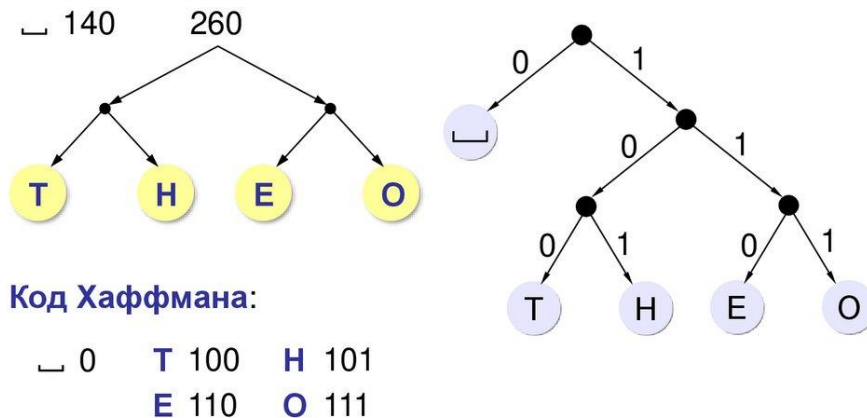


Задание 4 : Кодирование и декодирование информации

Алгоритм Хаффмана



Код Шеннона-Фано

Алгоритм Шеннона — Фано — один из первых алгоритмов сжатия.

Алгоритм префиксные, то есть никакое кодовое слово не является началом любого другого. Это свойство позволяет однозначно декодировать любую последовательность кодовых слов.

Выполнила: ст.гр.АБ-46 Федонина Алёна Олеговна

Условие Фано: ни одно кодовое слово не должно являться началом другого кодового слова (что обеспечивает однозначное декодирование сообщений с начала)

Префиксный код — это код, в котором ни одно кодовое слово не совпадает с началом другого кодового слова. Сообщения при использовании такого кода декодируются однозначно.

Обратное условие Фано: никакое кодовое слово не является окончанием другого кодового слова

Задание 1:

Для 5 букв латинского алфавита заданы их двоичные коды (для некоторых букв — из двух бит, для некоторых — из трех). Эти коды представлены в таблице:

a	b	c	d	e
000	110	01	001	10

Какой набор букв закодирован двоичной строкой 1100000100110?

Решение.

Мы видим, что выполняется условие Фано: никакое кодовое слово не является началом другого кодового слова, поэтому однозначно можем декодировать сообщение с начала.

Разобьём код слева направо по данным таблицы и переведём его в буквы:
110 000 01 001 10 — b a c d e.

Задание 2:

Для кодирования некоторой последовательности, состоящей из букв А, Б, В, Г и Д, используется неравномерный двоичный код, позволяющий однозначно декодировать полученную двоичную последовательность. Вот этот код: А — 1; Б — 0100; В — 000; Г — 011; Д — 0101. Требуется сократить для одной из букв длину кодового слова так, чтобы код по-прежнему можно было декодировать однозначно. Коды остальных букв меняться не должны. Каким из указанных способов это можно сделать?

- 1) для буквы Г — 11
- 2) для буквы В — 00
- 3) для буквы Г — 01
- 4) это невозможно

Решение.

Для однозначного декодирования получившееся в результате сокращения кодовое слово не должно быть началом никакого другого. Первый вариант ответа не подходит, поскольку код буквы А является началом кода буквы Г. Второй вариант ответа подходит. Третий вариант ответа не подходит, т. к. в таком случае код буквы Г является началом кода буквы Д.

Правильный ответ указан под номером: 2.

Задание 3:

По каналу связи с помощью равномерного двоичного кода передаются сообщения, содержащие только 4 буквы П, Р, С, Т. Каждой букве соответствует своё кодовое слово, при этом для набора кодовых слов выполнено такое свойство:

любые два слова из набора отличаются не менее чем в трёх позициях.

Это свойство важно для расшифровки сообщений при наличии помех. Для кодирования букв П, Р, С используются 5-битовые кодовые слова: П: 01111, Р: 00001, С: 11000. 5-битовый код для буквы Т начинается с 1 и заканчивается на 0. Определите кодовое слово для буквы Т.

Решение.

Код Т начинается с 1 и заканчивается на 0. Код С также начинается с 1 и заканчивается на 0. Поэтому для того, чтобы коды отличались не менее чем в трёх позициях, нужно, чтобы в остальных позициях все цифры были разные. И раз у С в середине 100, то у Т должно быть 011. Итого получили код 10110.

Ответ: 10110.

Задание 4:

По каналу связи передаются сообщения, содержащие только семь букв: А, Б, Г, И, М, Р, Я. Для передачи используется двоичный код, удовлетворяющий условию Фано. Кодовые слова для некоторых букв известны: А — 010, Б — 011, Г — 100. Какое **наименьшее** количество двоичных знаков потребуется для кодирования слова МАГИЯ?

Примечание. Условие Фано означает, что ни одно кодовое слово не является началом другого кодового слова.

Решение.

Следующая буква должна кодироваться как 11, поскольку 10 мы взять не можем. 100 взять не можем из-за Г, значит, следующая буква должна быть закодирована кодом 101. Следующая буква должна кодироваться как 000, поскольку 00 взять не можем, иначе не останется кодовых слов для оставшейся буквы, которые удовлетворяют условию Фано. Значит, последняя буква будет кодироваться как 001. Тогда наименьшее количество двоичных знаков, которые потребуются для кодирования слова МАГИЯ равно $2 + 3 + 3 + 3 + 3 = 14$.

Ответ: 14.

Задание 5:

По каналу связи передаются сообщения, содержащие только семь букв: А, Б, Г, И, М, Р, Я. Для передачи используется двоичный код, удовлетворяющий условию Фано. Кодовые слова для некоторых букв известны: А — 010, Б — 011, И — 10. Какое **наименьшее** количество двоичных знаков потребуется для кодирования слова ГРАММ?

Примечание. Условие Фано означает, что ни одно кодовое слово не является началом другого кодового слова.

Решение.

Для трёх букв кодовые слова уже известны, осталось подобрать для оставшихся четырёх букв такие кодовые слова, которые обеспечат наименьшее количество двоичных знаков для кодирования слова ГРАММ.

Закодируем букву М кодовым словом 00, поскольку буква М повторяется в слове ГРАММ два раза. Для буквы Г возьмём кодовое слово 110. Кодовое слово 111 взять не можем, поскольку для остальных букв не останется кодовых слов, удовлетворяющих условию Фано. Оставшиеся две буквы закодируем кодовыми словами длины 4.

Таким образом, наименьшее количество двоичных знаков, которые потребуются для кодирования слова ГРАММ, равно $3 + 4 + 3 + 2 + 2 = 14$.

Ответ: 14.

Задание 6:

По каналу связи передаются сообщения, содержащие только семь букв: А, Б, В, Г, Й, К, Л. Для передачи используется двоичный код, удовлетворяющий условию Фано. Кодовые слова для некоторых букв известны: Б — 00, Г — 010, К — 101. Какое наименьшее количество двоичных знаков потребуется для кодирования слова БАЛАЛАЙКА?

Примечание. Условие Фано означает, что ни одно кодовое слово не является началом другого кодового слова.

Решение.

Букву А закодируем кодовым словом 11, поскольку буква А повторяется в слове БАЛАЛАЙКА 4 раза. Букву Л закодируем кодовым словом 011, поскольку буква Л повторяется в слове БАЛАЛАЙКА 2 раза. Буквы Й и В закодируем кодовыми словами 1000 и 1001 соответственно. Тогда наименьшее количество двоичных знаков, которые потребуются для кодирования слова БАЛАЛАЙКА равно $2 + 2 + 3 + 2 + 3 + 2 + 4 + 3 + 2 = 23$.

Ответ: 23.

Задание 7:

Для передачи чисел по каналу с помехами используется код проверки четности. Каждая его цифра записывается в двоичном представлении, с добавлением ведущих нулей до длины 4, и к получившейся последовательности дописывается сумма её элементов по модулю 2 (например, если передаём 23, то получим последовательность 0010100110). Определите, какое число передавалось по каналу в виде 01100010100100100110.

Решение:

- Рассмотрим пример из условия задачи:

Было 23_{10}

Стало 0010100110_2

- Где сами цифры исходного числа (выделим их красным цветом):

0010100110 (0010 - 2, 0011 - 3)

- Первая добавленная цифра 1 после двоичной двойки — это проверка четности (1 единица в 0010 — значит нечетное), 0 после двоичной тройки — это также проверка нечетности (2 единицы в 0011, значит — четное).
- Исходя из разбора примера решаем нашу задачу так: поскольку «нужные» нам цифры образуются из групп по 4 числа в каждой плюс одно число на проверку четности, то разобьем закодированное сообщение на группы по 5, и отбросим из каждой группы последний символ:

- разбиваем по 5:

01100 01010 01001 00110

- отбрасываем из каждой группы последний символ:

0110 0101 0100 0011

- Результат переводим в десятичную систему:

0110 0101 0100 0011

↓ ↓ ↓ ↓

Ответ: 6 5 4 3

Перевод из двоичной в восьмеричную и шестнадцатеричную системы

$$a_1 a_2 \dots a_{n-2} a_{n-1} a_n$$

Для восьмеричной

000	0
001	1
010	2
011	3
100	4
101	5
110	6
111	7

$$a_1 a_2 \dots a_{n-4} a_{n-3} a_{n-2} a_{n-1} a_n$$

Для шестнадцатеричной

0000	0	1000	8
0001	1	1001	9
0010	2	1010	A
0011	3	1011	B
0100	4	1100	C
0101	5	1101	D
0110	6	1110	E
0111	7	1111	F

Для кодирования букв О, К, Г, Д, Р решили использовать двоичное представление чисел 0, 1, 2, 3 и 4 соответственно (с сохранением одного незначащего нуля в случае одноразрядного представления). Закодируйте последовательность букв ГОРОДОК таким способом и результат запишите восьмеричным кодом.

Решение.

Сначала следует представить данные в условии числа в двоичном коде:

О	К	Г	Д	Р
0	1	2	3	4
00	01	10	11	100

Затем закодировать последовательность букв: ГОРОДОК — 100010000110001. Теперь разобьём это представление на тройки справа налево и переведём полученный набор чисел в десятичный код, затем в восьмеричный (восьмеричное представление совпадает с десятичным при разбиении тройками)

100 010 000 110 001 — 42061.

Для кодирования букв О, Л, А, З, К решили использовать двоичное представление чисел 0, 1, 2, 3 и 4 соответственно (с сохранением одного незначащего нуля в случае одноразрядного представления). Закодируйте последовательность букв ЗАКОЛКА таким способом и результат запишите шестнадцатеричным кодом.

Решение.

Сначала следует представить данные в условии числа в двоичном коде:

О	Л	А	З	К
---	---	---	---	---

0	1	2	3	4
00	01	10	11	100

Затем закодировать последовательность букв: ЗАКОЛКА — 1110100000110010. Теперь разобьём это представление на четвёрки справа налево и переведем полученный набор чисел сначала в десятичный код, затем в шестнадцатеричный:

1110 1000 0011 0010 — 14 8 3 2 — E832.

Для передачи по каналу связи сообщения, состоящего только из символов А, Б, В и Г, используется неравномерный (по длине) код: А-10, Б-11, В-110, Г-0. Через канал связи передаётся сообщение: ВАГБААГВ. Закодируйте сообщение данным кодом. Полученное двоичное число переведите в восьмеричный вид.

Решение.

Закодируем последовательность букв: ВАГБААГВ — 1101001110100110. Теперь разобьём это представление на тройки справа налево и переведем полученный набор чисел сначала в десятичный код, затем в восьмеричный (в данном случае они совпадают):

1 101 001 110 100 110 для самой левой цифры 1 допишем два нуля слева, тогда получим

001 101 001 110 100 110 — 1 5 1 6 4 6.