

## § 2. Кодирование информации. Основные понятия

**Кодом** называют правило (или совокупность правил), в соответствии с которым производится отображение дискретных сообщений сигналами в виде определённых сочетаний символов вторичного алфавита.

**Кодирование** — это перевод информации, представленной символами первичного алфавита, в последовательность кодов.

**Декодирование** (операция, обратная кодированию) — перевод последовательности кодов в соответствующий набор символов первичного алфавита.

Операции кодирования и декодирования называются *обратимыми*, если их последовательное применение не приводит к потере информации.

Код является *однозначно декодируемым*, если любое слово, составленное из кодовых слов, можно декодировать только единственным способом.

По условию построения кодовых комбинаций коды делят на *равномерные* и *неравномерные*.

В равномерных кодах все сообщения передаются кодовыми группами с одинаковым числом элементов.

*Равномерное кодирование всегда допускает однозначное декодирование.*

**Пример 2.1.** Вася и Петя играли в шпионов и кодировали сообщения собственным шифром. Фрагмент кодовой таблицы приведён ниже:

К	Р	З	А	Г	О
#-	!-	-#	#!-	-	!

Расшифруйте сообщение, если известно, что буквы в нём не повторяются:  
-!-!-##!-

Запишите в ответе расшифрованное сообщение.

*Решение.* Первый символ в сообщении -. Из таблицы находим, что с этого символа начинается только код буквы З и этот символ соответствует коду буквы Г. Так как в шифровке за - следует символ !, то код буквы З не подходит. Следовательно Г — первая буква зашифрованного сообщения.

Далее будем расшифровывать сообщение !-!-##!-. Первый символ этого сообщения — !. С этого символа начинается только код буквы Р и этот символ соответствует коду буквы О. Так как в шифровке за ! следуют символы -!, то код буквы О не подходит. В противном случае в расшифрованном сообщении будут повторяться буквы. Следовательно Р — вторая буква зашифрованного сообщения.

Расшифруем оставшуюся последовательность символов !-##!-.

Чтобы не было повторений букв в расшифровке, третьим символом должна быть буква О. Далее, нетрудно заметить, должна следовать последовательность букв ЗА.

Таким образом, расшифрованное сообщение имеет вид ГРОЗА.

*Ответ:* ГРОЗА.

**Пример 2.2.** Ваня шифрует русские слова, записывая вместо каждой буквы её номер в алфавите (без пробелов). Номера букв даны в таблице.

А	1	Е	6	Й	11	О	16	У	21	Ш	26	Э	31
Б	2	Ё	7	К	12	П	17	Ф	22	Щ	27	Ю	32
В	3	Ж	8	Л	13	Р	18	Х	23	Ъ	28	Я	33
Г	4	З	9	М	14	С	19	Ц	24	Ы	29		
Д	5	И	10	Н	15	Т	20	Ч	25	Ь	30		

Некоторые шифровки можно расшифровать не одним способом. Например, 12181620 может означать «КРОТ», может — «АУЖАЕТ», а может — «КАЖОТ». Даны четыре шифровки:

5212  
4620  
61920  
53212

Только одна из них расшифровывается единственным способом. Найдите её и расшифруйте. Результат расшифровки запишите в качестве ответа.

*Решение.* Первая шифровка 5212 может быть расшифрована тремя способами:

1) ДБАБ (5 — Д, 2 — Б, 1 — А, 2 — Б), 2) ДУБ (5 — Д, 21 — У, 2 — Б), 3) ДБК (5 — Д, 2 — Б, 12 — К).

Вторая шифровка 4620 может быть расшифрована только одним способом: ГЕТ (4 — Г, 6 — Е, 20 — Т).

Третья шифровка 61920 может быть расшифрована двумя способами: 1) ЕАЗТ (6 — Е, 1 — А, 9 — З, 20 — Т), 2) ЕСТ (6 — Е, 19 — С, 20 — Т).

Четвёртая шифровка 53212 может быть расшифрована пятью способами: 1) ДВБАБ (5 — Д, 3 — В, 2 — Б, 1 — А, 2 — Б), 2) ДЮАБ (5 — Д, 32 — Ю, 1 — А, 2 — Б), 3) ДЮК (5 — Д, 32 — Ю, 12 — К), 4) ДВУБ (5 — Д, 3 — В, 21 — А, 2 — Б), 5) ДВБК (5 — Д, 3 — В, 2 — Б, 12 — К).

Следовательно, искомая шифровка 4620, результатом её расшифровки является ГЕТ.

*Ответ:* ГЕТ.

**Пример 2.3.** Для кодирования букв *Г, Н, О, Ь* использовали четырёхразрядные двоичные числа, начинающиеся и оканчивающиеся на 1 (от 1001 до 1111 соответственно). В результате кодирования некоторого сообщения получили последовательность 11011001110110111111.

Определите закодированное сообщение.

*Решение.* Из условия следует, что заданные буквы имеют следующие коды:

*Г* — 1001, *Н* — 1011, *О* — 1101, *Ь* — 1111.

Для расшифровки достаточно разбить кодовую последовательность 11011001110110111111 на четвёрки: 1101 1001 1101 1011 1111 — и на месте каждой из них записать соответствующую букву.

Получим сообщение *ОГОНЬ*.

*Ответ:* ОГОНЬ.

## § 3. Дискретное (цифровое) представление информации

### 3.1. Единицы измерения информации и их производные

**Бит** — минимальная единица количества информации, равная одному двоичному разряду.

**Байт** — единица количества информации, являющаяся наименьшей единицей памяти компьютера и равная 8 битам.

Для больших объёмов информации используют производные единицы измерения:

1 б (байт) = 8 бит (8 двоичных разрядов)

1 Кб (килобайт) =  $2^{10}$  б = 1024 б

1 Мб (мегабайт) =  $2^{20}$  б = 1024 Кб

1 Гб (гигабайт) =  $2^{30}$  б = 1024 Мб

1 Тб (терабайт) =  $2^{40}$  б = 1024 Гб

1 Пб (петабайт) =  $2^{50}$  б = 1024 Тб

### 3.2. Информационный объём

**Информационным объёмом сообщения** называется количество двоичных символов, которое используется для кодирования этого сообщения.

### 3.3. Представление текстовой информации.

#### Основные кодировки

Если с каждым символом алфавита сопоставить определённое целое число (например, порядковый номер), то с помощью двоичного кода можно кодировать текстовую информацию.

Основные кодировки.

1. ASCII (American Standard Code for Information Interchange — стандартный код информационного обмена США). ASCII представляет собой 8-битную кодировку для представления десятичных цифр, латинского и национального алфавитов, знаков препинания и управляющих символов. Нижнюю половину кодовой таблицы (0 — 127) занимают символы US-ASCII (см. приложение на стр. 166—174), а верхнюю (128 — 255) — дополнительные символы, включая набор национальных символов.

2. Windows-1251 — кодировка символов русского языка; используется на платформе Windows. Каждому символу в кодировке Windows-1251 соответствует 8-битовый двоичный код.

3. КОИ-8 (код обмена информацией, восьмизначный); встречается в компьютерных сетях на территории России и в российском секторе Интернета.

4. Unicode — система, основанная на 16-разрядном кодировании символов. Шестнадцать разрядов позволяют обеспечить уникальные коды для 65536 различных символов. Каждому символу в этой кодировке соответствует 16-битовый (2-байтовый) двоичный код. Первые 128 символов Unicode совпадают с соответствующими символами ASCII.

**Пример 3.1.** Определите количество информационного объёма выражения «Жёсткий диск», записанного в кодировке Unicode.

*Решение.* В системе Unicode каждый символ кодируется двумя байтами. В приведённом выражении 12 символов. Следовательно, информационный объём этого выражения равен  $12 \cdot 2 = 24$  байта.

*Ответ:* 24 байта.

**Пример 3.2.** Автоматическое устройство осуществило перекодировку информационного сообщения длиной в 16 символов, записанного на русском языке в 8-битном коде КОИ-8, в 16-битную кодировку Unicode. На сколько бит при этом увеличился информационный объём сообщения?

*Решение.* В 16-битном коде на 1 символ отводится на 8 бит больше, чем в 8-битной кодировке. Следовательно, информационный объём сообщения длиной в 16 символов увеличился на  $8 \cdot 16 = 128$  бит.

*Ответ:* 128 бит.

**Пример 3.3.** Рассказ занимает на жёстком диске 60 Кб. На одной странице 40 строк по 32 символа в строке, каждый символ кодируется 16 битами в представлении Unicode. Сколько страниц содержит рассказ?

*Решение.* На одной странице содержится  $40 \cdot 32 = 1280$  символов. Так как один символ кодируется 16 битами, то информационный объём одной страницы в кодировке Unicode  $1280 \cdot 16 = 20\,480$  бит  $= 2\,560$  байт  $= 2,5$  Кб. Рассказ занимает 60 Кб, значит он содержит  $60/2,5 = 24$  страницы.

*Ответ:* 24.

### 3.4. Представление числовой информации

**Представление целых неотрицательных чисел.** Если для представления целого числа в памяти компьютера отведено  $N$  бит, то количество различных значений будет равно  $2^N$ .

Для кодирования целых чисел от 0 до 255 достаточно иметь 8 разрядов двоичного кода (8 бит).

Для кодирования целых чисел от 0 до 65 535 требуется шестнадцать бит; 24 бита позволяют закодировать более 16,5 миллионов разных значений.

Максимальное значение целого неотрицательного числа достигается в случае, когда во всех ячейках стоят единицы. Если под представление целого положительного числа отведено  $N$  бит, то максимальное значение будет равно  $2^N - 1$ .

**Пример 3.4.** 1) Сколько различных значений можно закодировать 8 битами? 2) Укажите наибольшее целое неотрицательное число, которое можно закодировать 8 битами (числа кодируются своим представлением в двоичной системе счисления).

*Решение.* 1) Количество различных значений, которое можно закодировать  $N$  битами равно  $2^N$ . Следовательно 8 битами можно закодировать  $2^8 = 256$  различных значений.

2) Максимальное значение целого неотрицательного числа достигается в случае, когда во всех ячейках стоят единицы. Если под представление целого положительного числа отведено  $N$  бит, то максимальное значение будет равно  $2^N - 1$ . Следовательно наибольшее целое неотрицательное число, которое можно закодировать 8 битами равно  $2^8 - 1 = 255$ .

*Ответ:* 1)256; 2)255.

### 3.5. Графическая информация. Основные понятия

**Пиксель** — это минимальный участок изображения, для которого независимым образом можно задать цвет.

**Глубина цвета** — количество информации, которое используется для кодирования цвета точки изображения. Наиболее распространёнными глубинами цвета являются 4, 8, 16 и 24 бита на точку.

Количество различных цветов  $N$  и количество информации  $I$ , необходимое для кодирования каждой точки, связаны формулой  $N = 2^I$ .

Глубина цвета, $I$ (бит)	Количество цветов в палитре, $N$	Глубина цвета, $I$ (бит)	Количество цветов в палитре, $N$
1	2	7	128
2	4	8	256
3	8	16 (High Color)	65536
4	16	24 (True color)	16777216
5	32	32 (True color)	4294967296
6	64		

**Разрешающая способность экрана** — количество пикселей на единицу длины, dpi (dots per inch — «точка на дюйм»).

### Расчёт объёма видеопамати.

Информационный объём требуемой для хранения изображения видеопамати можно рассчитать по формуле:

$$I_{\Pi} = I \cdot X \cdot Y,$$

где  $I_{\Pi}$  — информационный объём видеопамати в битах;

$X \cdot Y$  — количество точек изображения ( $X$  — количество точек по горизонтали,  $Y$  — по вертикали);

$I$  — глубина цвета в битах на точку.

**Пример 3.5.** Определите необходимый объём видеопамати для графического режима с пространственным разрешением  $1024 \times 768$  точек и глубиной цвета 24 бита.

*Решение.*  $I_{\Pi} = I \cdot X \cdot Y = 24 \text{ бита} \cdot 1024 \cdot 768 = 18\,874\,368 \text{ бит} = 2\,359\,296 \text{ байт} = 2\,304 \text{ Кбайт} = 2,25 \text{ Мбайт}.$

*Ответ:* 2,25 Мбайт.

### 3.6. Представление звуковой информации

Звук представляет собой распространяющуюся в воздухе, воде или другой среде волну с непрерывно меняющейся интенсивностью и частотой.

Для того чтобы компьютер мог обрабатывать звук, непрерывный звуковой сигнал должен быть преобразован в цифровую дискретную форму с помощью *временной дискретизации*.

**Частота дискретизации звука** — это количество измерений уровня звукового сигнала за одну секунду.

Частота дискретизации звука может лежать в диапазоне от 8000 до 48 000 измерений уровня звукового сигнала за одну секунду.

Уровни звукового сигнала можно рассматривать как набор возможных состояний  $N$ , для кодирования которых необходимо определённое количество информации  $I$ , которое называется глубиной кодирования звука.

**Глубина кодирования звука** — это количество информации (количество бит), которое необходимо для кодирования дискретных уровней громкости цифрового звука.

Уровень цифрового звукового сигнала можно рассчитать по формуле:  $N = 2^I$ .

**Пример 3.13.** Пусть глубина кодирования звука составляет 16 битов, тогда количество уровней громкости звука:  $N = 2^I = 2^{16} = 65536$ .

**Размер (в байтах) цифрового моноаудиофайла** можно оценить по формуле:  $A = D \cdot T \cdot I/8$ ,

где  $D$  — частота дискретизации звука за одну секунду,  $T$  — время звучания или записи звука (с),  $I$  — глубина кодирования звука (бит).

**Пример 3.6** Производится четырёхканальная (квадро) звукозапись с частотой дискретизации 64 кГц и 16-битной глубиной кодирования. Запись длится 3 минуты, её результаты записываются в файл, сжатие данных не производится. Определить размер полученного файла.

*Решение.* Согласно условию задачи  $D = 64000$  (изм. в с),  $T = 3$  мин = 180 с,  $I = 16$  бит. Следовательно,  
 $A = D \cdot T \cdot I/8 = 64000 \cdot 180 \cdot 16/8 = 23040000$  б  $\approx 23$  Мб. При четырёхканальной (квадро) записи объём памяти, необходимый для хранения данных одного канала, умножается на 4;  $23$  Мб  $\cdot 4 = 92$  Мб.

*Ответ:* 92 Мб.