

# Представление и измерение информации

Представление текстовой информации

Представление числовой информации

Представление графической информации

Представление звуковой информации

Измерение информации

# Числовая форма

## представления

Компьютер может обрабатывать числовую, текстовую, графическую, звуковую, видеоинформацию только тогда, когда она представлена в нем в двоичном коде, т. е. используется алфавит мощностью в два символа: 0 и 1.

**Бит** – это единица информации, представляющая собой двоичный разряд, который может принимать значение 0 или 1.

**Байт** – это восемь последовательных бит. В 1 байте можно кодировать значение одного символа из 256 возможных комбинаций.

1 Кбайт = 1024 байт; 1 Мбайт = 1024 Кбайт; 1 Гбайт = 1024 Мбайт и так далее.

# Представление текстовой информации

Чтобы закодировать один символ, традиционно используют количество информации, равное 1 байту, т. е.  $1 = 1 \text{ байт} = 8 \text{ бит}$ . В 60-е годы XX века это было закреплено комитетом ASCII США в ASCII-стандарте.

Суть кодирования заключается в том, что каждому символу ставят в соответствие двоичный код от 00000000 до 11111111 или соответствующий ему десятичный код от 0 до 255.

В середине 90-х годов XX века появилась новая кодировка – Unicode, поддерживающая 65 536 различных символов. В ней на каждый символ отводится по 2 байта:

$$K = 2^{16} = 65\,536.$$

32	20		56	38	8	80	50	P	104	68	h
33	21	!	57	39	9	81	51	Q	105	69	i
34	22	“	58	3A	:	82	52	R	106	6A	j
35	23	#	59	3B	;	83	53	S	107	6B	k
36	24	\$	60	3C	<	84	54	T	108	6C	l
37	25	%	61	3D	=	85	55	U	109	6D	m
38	26	&	62	3E	>	86	56	V	110	6E	n
39	27	‘	63	3F	?	87	57	W	111	6F	o
40	28	(	64	40	@	88	58	X	112	70	p
41	29	)	65	41	A	89	59	Y	113	71	q
42	2A	*	66	42	B	90	5A	Z	114	72	r
43	2B	+	67	43	C	91	5B	[	115	73	s
44	2C	,	68	44	D	92	5C	\	116	74	t
45	2D	-	69	45	E	93	5D	]	117	75	u
46	2E	.	70	46	F	94	5E	^	118	76	v
47	2F	/	71	47	G	95	5F	_	119	77	w
48	30	0	72	48	H	96	60	‘	120	78	x
49	31	1	73	49	I	97	61	a	121	79	y
50	32	2	74	4A	J	98	62	b	122	7A	z
51	33	3	75	4B	K	99	63	c	123	7B	{
52	34	4	76	4C	L	100	64	d	124	7C	
53	35	5	77	4D	M	101	65	e	125	7D	}
54	36	6	78	4E	N	102	66	f	126	7E	~
55	37	7	79	4F	O	103	67	g	127	7F	

# Представление числовой информации

Нормализованная запись числа

$$A = m \cdot q^p,$$

где  $m$  – мантисса числа  $A$ ;

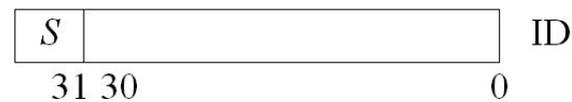
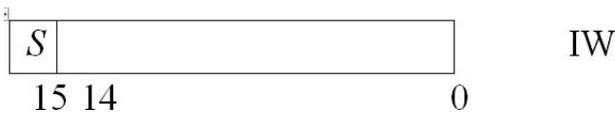
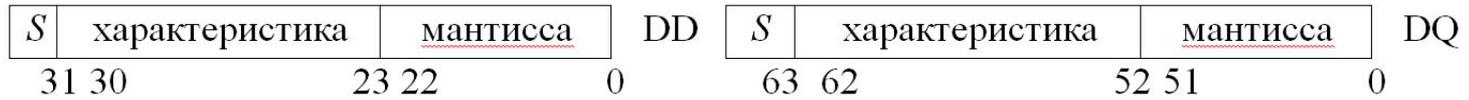
$p$  – порядок числа  $A$ ;

$q$  – основание системы счисления.

$$q^{-1} \leq |m_A| < 1$$

# Форматы чисел в памяти ПК

Формат данных	Число байт	Число значащих десятичных цифр	Диапазон значений
короткое вещественное DD	4	6–7	$\pm 1.2 \times 10^{-38} \div 3.4 \times 10^{38}$
длинное вещественное DQ	8	15–16	$\pm 2.3 \times 10^{-308} \div 1.8 \times 10^{308}$
расширенное вещественное DT	10	19–20	$\pm 3.4 \times 10^{-4932} \div 1.2 \times 10^{4932}$
целое слово IW	2	4–5	$(-32768) \div (+32767)$
короткое целое ID	4	9	$-2^{31} \div 2^{31}-1$
длинное целое IQ	8	18	$-2^{63} \div 2^{63}-1$



Вещественные числа в каждом из трех форматов имеют три поля: поле знака мантиссы  $S$ , поле порядка и поле мантиссы. Мантисса числа записывается в нормализованном виде

$$1.m_1m_2 \dots m_i.$$

Целая часть, всегда равная 1, прямо не представляется в форматах DD и DQ, а учитывается неявно. В формате DT старший бит мантиссы представляется явно.

Порядок вещественных чисел  $E$  записывается в поле порядка в смещенном виде. Он равен истинному порядку, увеличенному на значение смещения

$$E = \text{истинный порядок} + \text{смещение}.$$

Значение смещения для соответствующих форматов равно 127, 1023 и 16383

$$(-1)^S \times (1.m_1m_2 \dots m_{23}) \times 2^{E-127} \quad \text{для DD};$$

$$(-1)^S \times (1.m_1m_2 \dots m_{52}) \times 2^{E-1023} \quad \text{для DQ};$$

$$(-1)^S \times (1.m_1m_2 \dots m_{64}) \times 2^{E-16383} \quad \text{для DT}.$$

# Представление целых чисел

положительного числа  $N$ :

- перевести число  $N$  в двоичную систему счисления;
- полученный результат дополнить слева незначащими нулями до полного заполнения формата целого числа.

отрицательного числа  $N$ :

- получить внутреннее представление целого положительного числа  $N$ ;
- получить обратный код этого числа заменой 0 на 1 и 1 на 0 (инвертирование);
- к полученному числу прибавить 1 – получается дополнительный код.

# Представление вещественных чисел

- перевести  $A_{10} \rightarrow A_2$ ;
- нормализовать число;
- выполнить смещение порядка;
- перевести смещенный порядок  $E_{10} \rightarrow E_2$ ;
- записать число в подходящем формате.

# Представление графической информации

Для представления графической информации в двоичной форме используются **растровый** и **векторный** способы.

При **растровом** способе вертикальными и горизонтальными линиями изображение разбивается на отдельные точки; каждому элементу ставятся в соответствие коды его цвета и место, которое он занимает.

В **векторном** способе информация вычисляется по специальным формулам, описывающим какой-либо объект – геометрический примитив (точка, прямая, дуга и другие).

Можно сочетать векторный и растровый способы формирования изображений (3D графика)

# Представление звуковой информации

**Семпл** – это промежуток времени между двумя измерениями амплитуды аналогового сигнала.

Процесс аналого-цифрового преобразования называют **семплированием** или **дискретизацией**.

Важными параметрами семплирования являются частота и разрядность.

**Частота** – это количество измерений амплитуды аналогового сигнала в секунду.

**Разрядность** указывает, с какой точностью происходят изменения амплитуды аналогового сигнала.

**Звук** – волновые колебания давления в упругой среде (воздухе, воде, металле и т.д.). Для обозначения звука часто используют термин «звуковая волна».

# Параметры звука

Основные параметры – частота и амплитуда колебаний.

**Частоту** звука измеряют в герцах (Гц – количество колебаний в секунду). Человеческое ухо способно воспринимать звук в широком диапазоне частот, от 16 Гц до 20 кГц.

**Амплитуду** звуковых колебаний называют звуковым давлением или силой звука, эта величина характеризует воспринимаемую громкость звука.

Сила звука измеряется в паскалях (Па), при этом самые слабые, едва различимые звуки имеют амплитуду 20 мкПа. Самые сильные звуки, ещё не выводящие слуховые органы из строя, могут иметь амплитуду до 200 Па (болевого порог). Из-за такого широкого диапазона значений (максимальное и минимальное значения отличаются на 6-7 порядков!) абсолютными величинами звукового давления пользоваться неудобно, на практике обычно используют логарифмическую шкалу децибелов.

Относительная сила звука  $L = 20 * \lg \frac{P_{зв}}{P_{пс}}$ ,

где  $P_{пс}$  – порог слышимости,  $P_{зв}$  – давление измеряемого звука.

**Звукозапись** – процесс сохранения информации о параметрах звуковых волн. Способы хранения или записи звука разделяются на аналоговые и цифровые.

При аналоговой записи на носителе размещается непрерывный «слепок» звуковой волны.

В компьютерах применяется исключительно цифровая форма записи звука.

При цифровой записи звук необходимо подвергнуть временной **дискретизации** и **квантованию**.

**Частота дискретизации звука** – это количество измерений громкости звука за одну секунду. Одно измерение в секунду соответствует частоте 1Гц, 1000 измерений в секунду – 1 кГц.

**Глубина кодирования звука** (измеряется в битах) – это количество информации, которое необходимо для кодирования одного значения громкости цифрового звука.

Формула для расчета размера цифрового моноаудиофайла

$$V = M * I * t,$$

где  $I$  – глубина звука (в битах);  $t$  – время звучания (в секундах);  $M$  – частота дискретизации (в Гц).

Если звук записывается для нескольких источников воспроизведения (колонки, колонки + сабвуфер и т.д.), то объем умножается на соответствующий множитель.

*Пример.* Оценить информационный объем цифрового стереозвукового файла длительность звучания 1 секунда при глубине кодирования звука 16 бит и частоте дискретизации 24 кГц.

$N = 24000 \text{ Гц} * 16 \text{ бит} * 2 (\text{стереозвук!}) * 1 \text{ сек} = 768000 \text{ бит} = 93,75 \text{ Кбайт}.$

# Измерение информации

**Объемный способ** измерения информации самый распространенный. Заключается в том, что объем информации равен количеству передаваемых символов. Так как любая информация в вычислительной технике закодирована в двоичной системе счисления, то введены следующие единицы измерения – бит и байт. Восемь битов составляют один байт.

**Энтропийный способ** измерения информации основан на том, что получатель информации имеет представления о возможных наступлениях некоторых событий, которые могут наступить с некоторой вероятностью. Общая мера неопределенности (энтропия) характеризуется некоторой математической зависимостью от совокупности этих вероятностей. Количество информации в данном случае определяется тем, насколько уменьшится эта мера после ее получения.

**Алгоритмический способ** измерения информации заключается в том, чтобы оценить сложность (размер) программы, созданной для воспроизведения того или иного сообщения.

# Энтропийный способ

В 1928 году Р. Хартли, а в 1948 году К. Шеннон предложили формулы для вычисления количества информации, основанные на энтропии.

Р. Хартли первым ввел в теорию передачи информации методологию «измерения количества информации», при это он считал, что информация – это «группа физических СИМВОЛОВ».

**Информационная энтропия** – неопределённость появления какого-либо символа первичного алфавита. При отсутствии информационных потерь численно равна количеству информации на символ передаваемого сообщения.

Информационная энтропия источника определяется соотношением

$$H(x) = -\sum_{i=1}^n p(i) \log_2 p(i),$$

где  $p(i)$  – вероятность появления в тексте символа  $i$ .

# Формула Хартли

Пусть передается последовательность из  $n$  символов

$$a_1 a_2 \dots a_n,$$

каждый из которых принадлежит алфавиту  $A_m$  ( $m$  – число символов в алфавите). Число различных вариантов таких последовательностей

$$K = m^n.$$

Тогда количество информации, содержащейся в такой последовательности вычисляется по формуле

$$I = \log_2 K.$$

Замечание 1. Хартли предполагал, что все символы алфавита могут с равной вероятностью встретиться в любом месте.

Замечание 2. Любое сообщение длины  $n$  в одном алфавите будет содержать одинаковое количество информации. То есть, в формуле Хартли не учитывается смысловое содержание информации.

# Формула Шеннона

Пусть передается последовательность из  $n$  СИМВОЛОВ

$$a_1 a_2 \dots a_n,$$

каждый из которых принадлежит алфавиту  $A_m$  ( $m$  – число символов в алфавите). Вероятность появления символа  $a_i$  в последовательности равна  $P_i$ . Тогда количество информации, содержащейся в последовательности вычисляется по формуле

$$I = -n \cdot \sum_{i=1}^m P_i \log_2 P_i.$$

При равновероятных событиях формула Шеннона переходит в формулу Хартли.