

**Подходы к
понятию
информации
и измерению.**

Основные вопросы:

- Основные подходы к определению понятия «**информация**»
- Виды и свойства информации.
- Кодирование информации. Языки кодирования.
- Представление информации в двоичной системе счисления. Количество и размер информации. Единицы измерения информации.

Термин “Информация”

происходит от латинского
слова **informatio** –
пояснение, разъяснение.



Информация - это сведения об объектах и явлениях окружающей среды, их параметрах, свойствах и состоянии, которые воспринимают информационные системы (живые организмы, управляющие машины и др.) в *процессе* жизнедеятельности и работы.

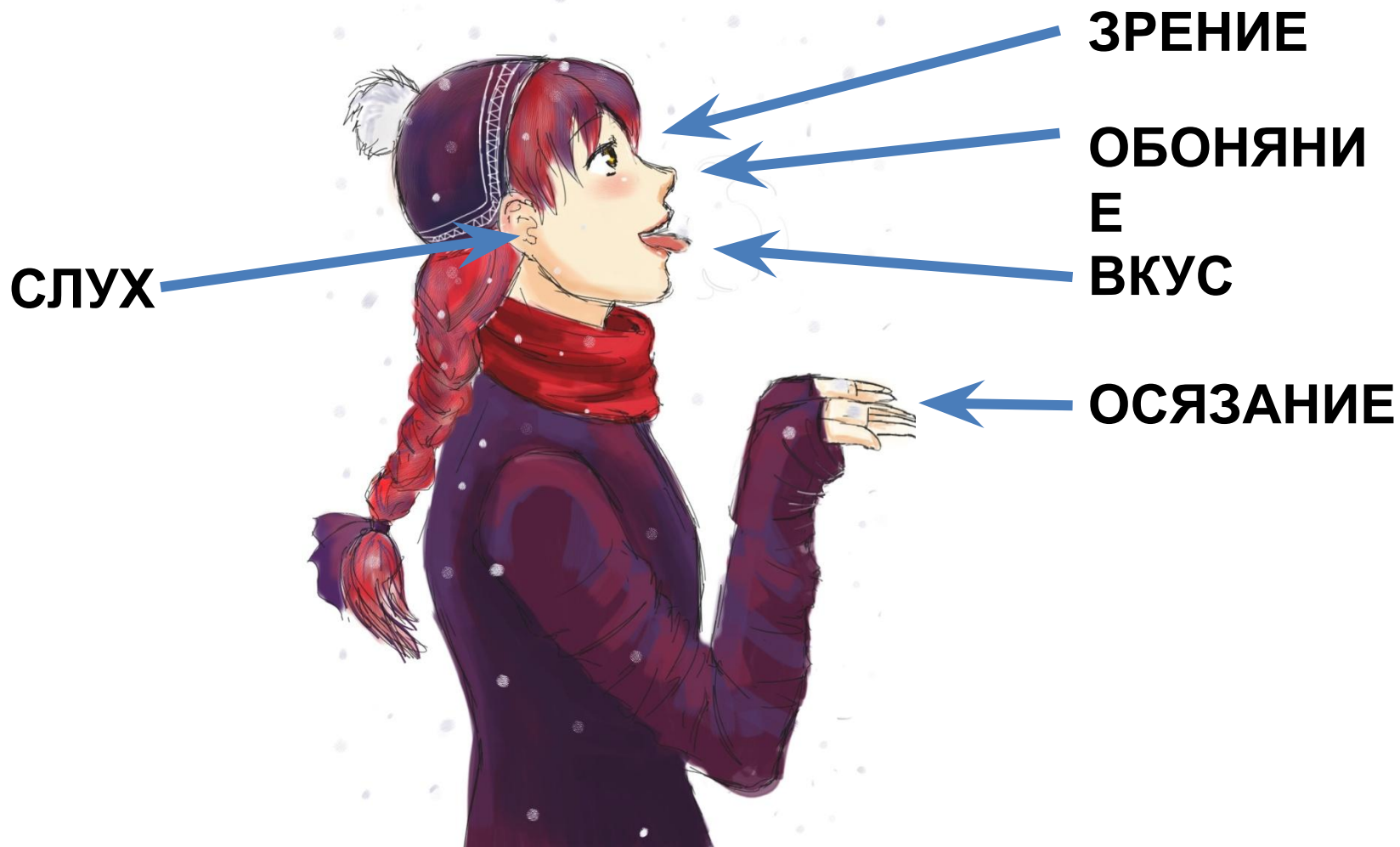


Информация – это ...

любые сведения об окружающем мире,
которые человек получает с помощью
органов чувств:

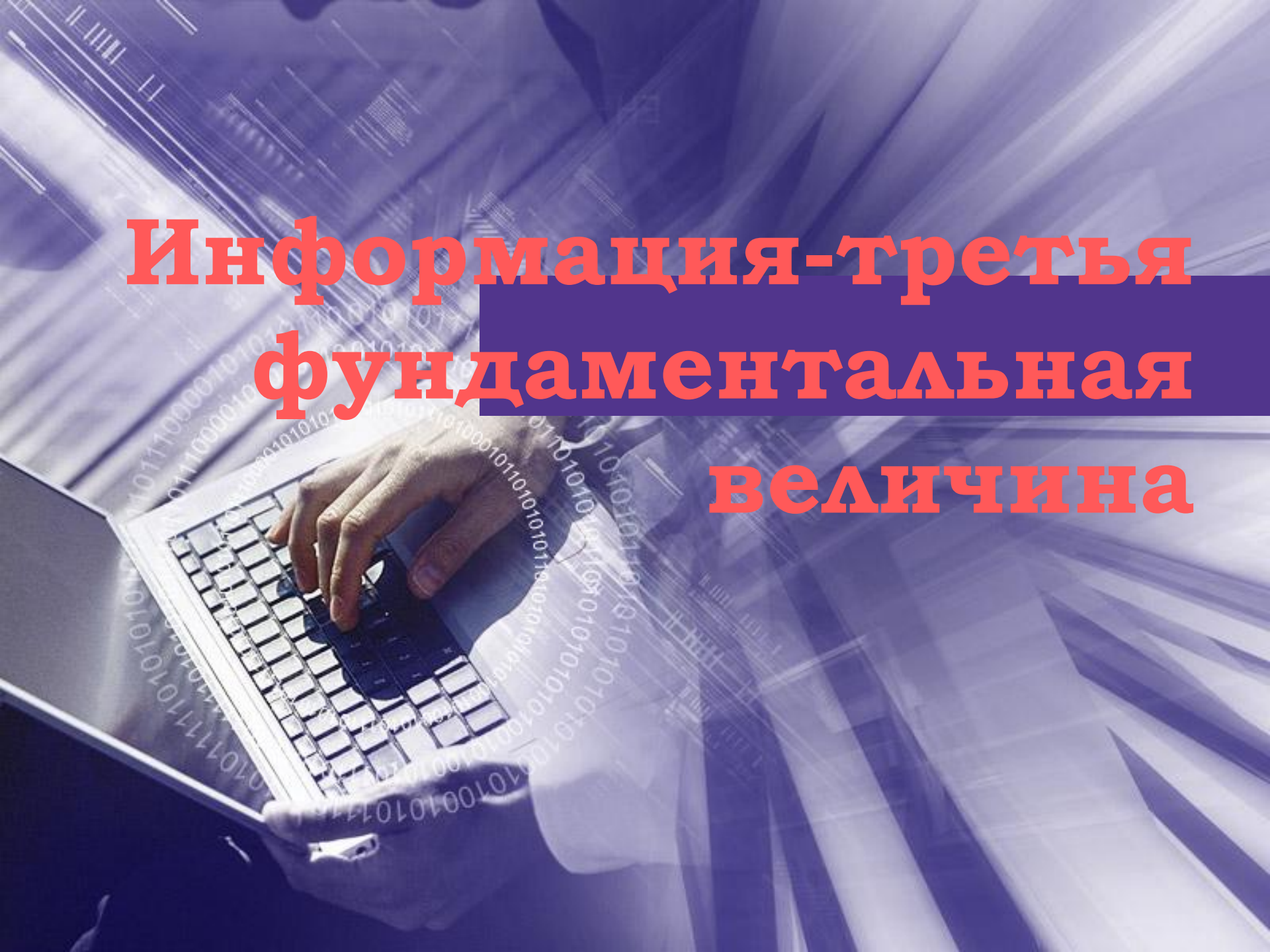
- глаза (зрение, 90 процентов информации)
- уши (слух)
- язык (вкус)
- нос (обоняние)
- кожа (осязание)

Человек получает информацию с помощью органов чувств



ИНФОРМАЦИЯ (ЛАТ. INFORMATIO) -
СВЕДЕНИЯ, РАЗЪЯСНЕНИЯ, ИЗЛОЖЕНИЕ.
ПОНЯТИЕ ИНФОРМАЦИИ





**Информация-третья
фундаментальная
величина**

ВИДЫ ИНФОРМАЦИИ

ПО СПОСОБУ ВОСПРИЯТИЯ

- визуальная
- аудиальная
- тактильная
- обонятельная
- вкусовая

ПО ФОРМЕ ПРЕДСТАВЛЕНИЯ

- текстовая;
- числовая;
- графическая;
- музыкальная;
- комбинированная

ПО ОБЩЕСТВЕННОМУ ЗНАЧЕНИЮ

- **массовая** (общенная, общественно-политическая, эстетическая)
- **специальная** (научная, производственная, техническая, управленческая)
- **личная** (знания, умения, интуиция)

По способу восприятия



Визуальная



Аудиальная



**Тактильная
(осязание)**



Вкусовая



**Обонятельна
я**

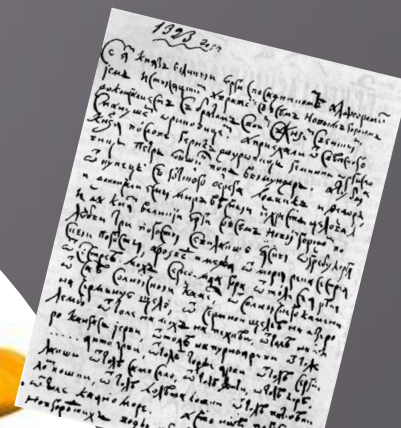
По форме представления



Звуковая



Графическая



Текстовая



Числовая

По назначению



Массовая - содержит обычные сведения и оперирует набором понятий, понятных большинству



Специальная - содержит специфический набор понятий, понятных в рамках узкой социальной группы



Личная (приватная) - набор сведений о какой-либо личности.



Секретная - передаваемая узкому кругу лиц и по закрытым (защищенным) каналам

Свойства информации:



Актуальная
(актуальность)

информация , ценная в данный
момент времени

Достоверная
(достоверность)

информация , полученная без
искажений

Полная (полнота)

информация , достаточная для
принятия правильного решения или
понимания

Понятная (понятность)


информация , выраженная на языке,
понятном тому, кому она
предназначена

Полезная (полезность)

Определяется субъектом, получившим
информацию в зависимости от
возможности ее использования

Точная (точность)

Определяется степенью ее близости к
реальному состоянию объекта,
процесса, явления и т. п.



*Кодирование информации.
Языки кодирования.
Универсальность
дискретного(цифрового)
представления
информации*

Для представления и обмена информацией между людьми служат ЯЗЫКИ



Естественные

Возникшие в результате исторического развития



Формальные

Созданные искусственно человеком для решения различных задач



$$Q = \frac{\pi}{4} \int_0^l d^2 dl.$$

А • —	Л • — • •	Ц — • — •
Б — • • •	М — —	Ч — — —
В • — —	Н — •	Ш — — —
Г — — •	О — — —	Щ — — • —
Д — • •	П • — • •	Ъ • — • — •
Е •	Р • • •	Ы — • — —
Ж • • • • —	С • • • •	Ь — • — —
З — — • •	Т — —	Э • • • • •
И • •	У • • —	Ю • • • —
Й • — —	Ф • • — •	Я • • — —
К • — —	Х • • • •	

Кодирование информации

– это запись информации с помощью некоторой знаковой системы (языка).

- **Код** – набор символов (условных обозначений) для представления информации.
- **Кодирование** – процесс представления информации (сообщения) в виде кода.
- **Декодирование** – процесс обратного преобразования кода к форме исходной символьной системы, т.е. получение исходного сообщения. *Например:* перевод с азбуки Морзе в письменный текст на русском языке.

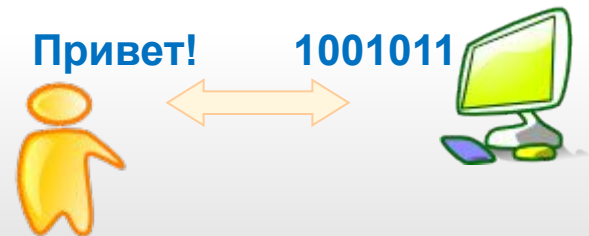
Двоичное кодирование в компьютере

Вся информация, которую обрабатывает компьютер должна

0 и

1.

Эти два символа принято называть двоичными цифрами или битами.



Кодирование символов

Текстовый файл

- на экране (символы)
- в памяти – двоичные коды



1000001 ₂	1000010 ₂	1000011 ₂	1000100 ₂
65	66	67	68



В файле хранятся не изображения символов, а их числовые коды в двоичной системе!



1 символ – 1 байт (8 бит)

Для кодирования **одного символа** требуется **один байт** информации.

Учитывая, что каждый **бит** принимает значение **1** или **0**, получаем, что с помощью 1 байта можно закодировать 256 различных **СИМВОЛОВ**.

$$2^8=256$$



Двоичное кодирование текстовой информации

Кодирование заключается в том, что
каждому символу ставится в
соответствие уникальный двоичный код
от 00000000 до 11111111 (или десятичный
код от 0 до 255).

Важно, что присвоение символу конкретного
кода – это вопрос соглашения, которое
фиксируется кодовой таблицей.



Таблица кодировки

Таблица, в которой всем символам компьютерного алфавита поставлены в соответствие порядковые номера (коды), называется таблицей кодировки.

Для разных типов ЭВМ используются различные кодировки. С распространением IBM PC международным стандартом стала таблица кодировки ASCII (American Standart Code for Information Interchange) – Американский стандартный код для информационного обмена.



Таблица кодировки ASCII

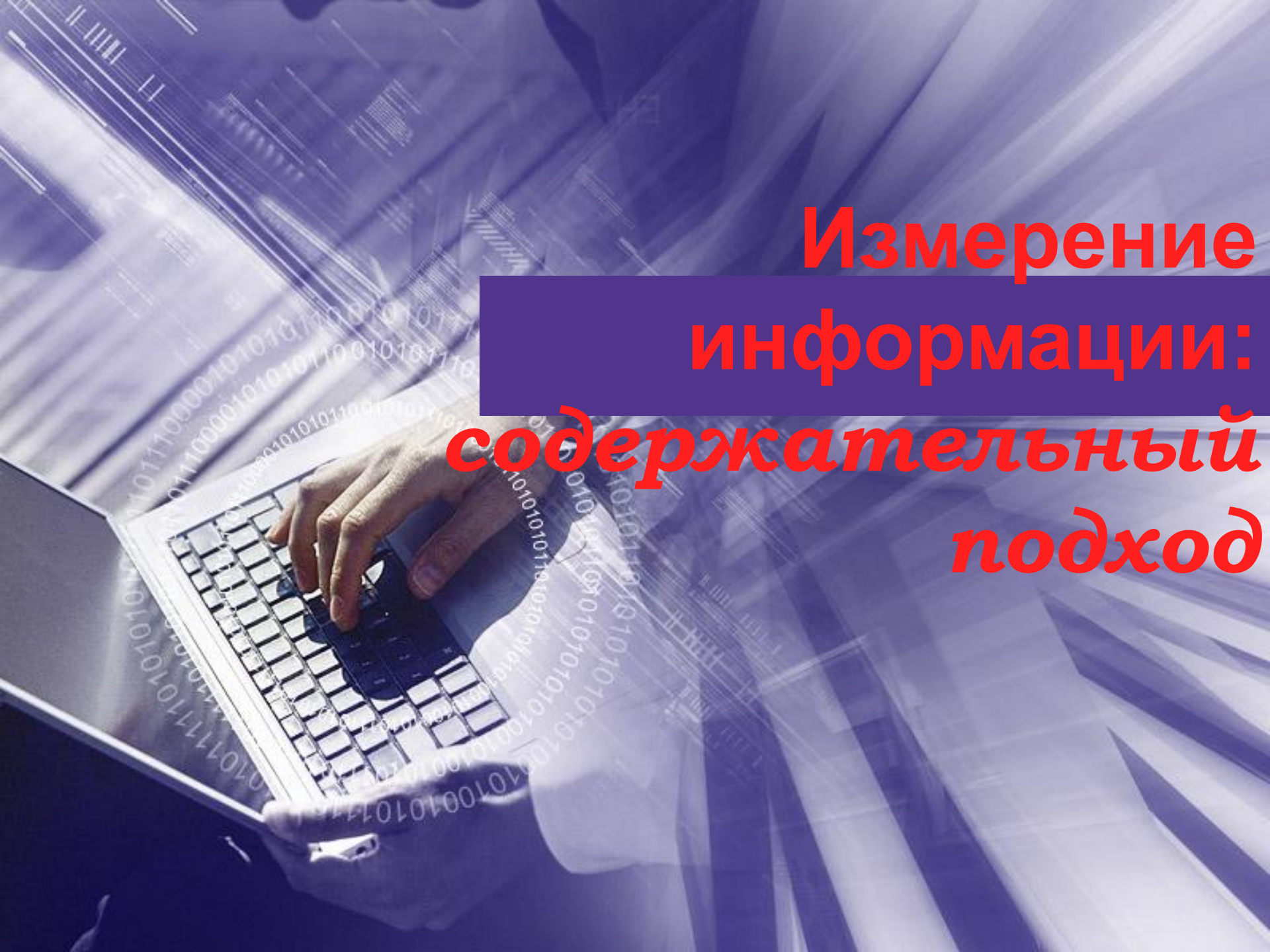
Стандартной в этой таблице является только первая половина, т.е. символы с номерами от 0 (00000000) до 127 (01111111). Сюда входят буква латинского алфавита, цифры, знаки препинания, скобки и некоторые другие символы.

Остальные 128 кодов используются в разных вариантах. В русских кодировках размещаются символы русского алфавита.

В настоящее время существует 5 разных кодовых таблиц для русских букв (KOI8, CP1251, CP866, Mac, ISO).

В настоящее время получил широкое распространение новый международный стандарт Unicode, который отводит на каждый символ два байта. С его помощью можно закодировать 65536 ($2^{16} = 65536$) различных символов.

символ	10-й код	2-й код	символ	10-й код	2-й код	символ	10-й код	2-й код	символ	10-й код	2-й код
	32	00100000	8	56	00111000	P	80	01010000	h	104	01101000
!	33	00100001	9	57	00111001	Q	81	01010001	i	105	01101001
"	34	00100010	:	58	00111010	R	82	01010010	j	106	01101010
#	35	00100011	;	59	00111011	S	83	01010011	k	107	01101011
\$	36	00100100	<	60	00111100	T	84	01010100	l	108	01101100
%	37	00100101	=	61	00111101	U	85	01010101	m	109	01101101
&	38	00100110	>	62	00111110	V	86	01010110	n	110	01101110
'	39	00100111	?	63	00111111	W	87	01010111	o	111	01101111
(40	00101000	@	64	01000000	X	88	01011000	p	112	01110000
)	41	00101001	A	65	01000001	Y	89	01011001	q	113	01110001
*	42	00101010	B	66	01000010	Z	90	01011010	r	114	01110010
+	43	00101011	C	67	01000011	[91	01011011	s	115	01110011
,	44	00101100	D	68	01000100	\	92	01011100	t	116	01110100
-	45	00101101	E	69	01000101]	93	01011101	u	117	01110101
.	46	00101110	F	70	01000110	^	94	01011110	v	118	01110110
/	47	00101111	G	71	01000111	_	95	01011111	w	119	01110111
0	48	00110000	H	72	01001000	`	96	01100000	x	120	01111000
1	49	00110001	I	73	01001001	a	97	01100001	y	121	01111001
2	50	00110010	J	74	01001010	b	98	01100010	z	122	01111010
3	51	00110011	K	75	01001011	c	99	01100011	{	123	01111011
4	52	00110100	L	76	01001100	d	100	01100100		124	01111100
5	53	00110101	M	77	01001101	e	101	01100101	}	125	01111101
6	54	00110110	N	78	01001110	f	102	01100110	~	126	01111110
7	55	00110111	O	79	01001111	g	103	01100111	□	127	01111111



**Измерение
информации:
содержательный
подход**

Для человека информация — это знания человека.

Получение новой информации приводит к расширению знаний.

Если некоторое сообщение приводит к уменьшению неопределенности нашего знания, то можно говорить, что такое сообщение содержит информацию.

Отсюда следует вывод, что сообщение информативно (т.е. содержит ненулевую информацию), если оно *пополняет знания человека.*

Нетрудно понять, что информативность одного и того же сообщения может быть разной для разных людей.

Например: « $2 \times 2 = 4$ » информативно для первоклассника, изучающего таблицу умножения, и неинформативно для старшеклассника.



Информативность сообщения

Но для того чтобы сообщение было информативно оно должно еще быть понятно.

Быть *понятным*, значит быть *логически связанным с предыдущими знаниями* человека.

Получение всяких знаний должно идти от простого к сложному. И тогда каждое новое сообщение будет в то же время понятным, а значит, будет нести информацию для человека.

Сообщение несет информацию для человека, если содержащиеся в нем сведения являются для него новыми и понятными.



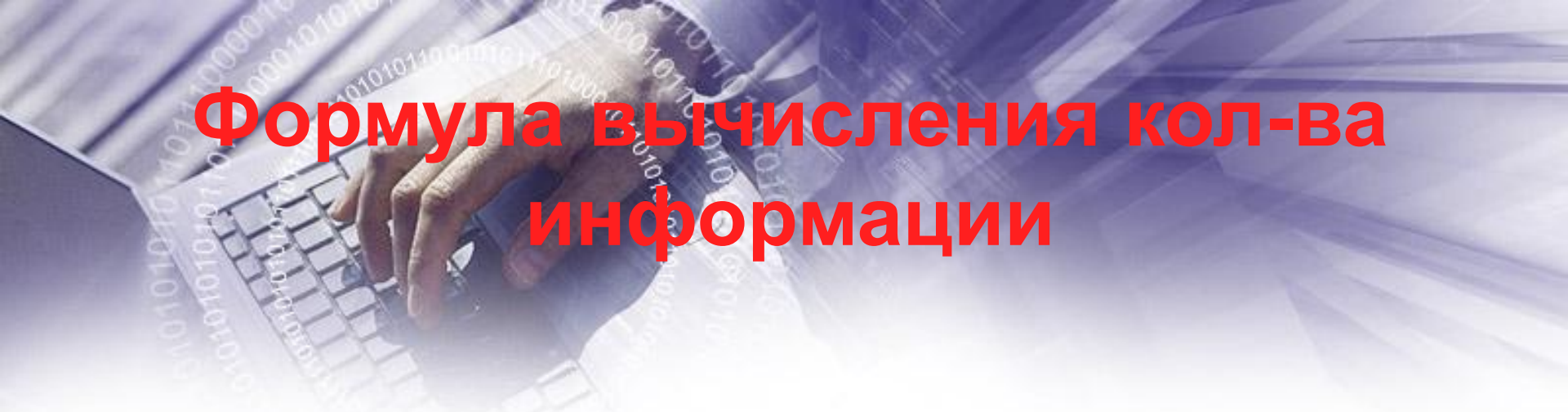
Единица измерения информации

Очевидно, различать лишь две ситуации: «нет информации» — «есть информация» для измерения информации недостаточно. Нужна единица измерения, тогда мы сможем определять, в каком сообщении информации больше, в каком — меньше.

Единица измерения информации была определена в науке, которая называется теорией информации. Эта единица носит название *«бит»*. Ее определение звучит так:

Сообщение, уменьшающее неопределенность знаний в два раза, несет 1 бит информации.

Неопределенность знаний о некотором событии — это количество возможных результатов события.



Формула вычисления кол-ва информации

Если обозначить *возможное количество событий*, или, другими словами, *неопределенность знаний* ***N***, а буквой ***I*** - *количество информации в сообщении* о том, что произошло одно из ***N*** событий, то можно записать формулу:

$$2^I = N$$



**Измерение
информации:
*алфавитный подход***



Способ измерения информации, который не связывает количество информации с содержанием сообщения, называется алфавитным подходом.

При алфавитном подходе к определению количества информации **отвлекаются от содержания информации** и рассматривают информационное сообщение как **последовательность знаков** определенной знаковой системы.

Алфавитный подход

Алфавит – набор знаков, используемых при кодировании информации с помощью некоторого языка.

Примеры:

АБВГДЕЖЗИЙКЛМНОПРСТУФХЦЧШЩЪЫЬЭЮЯ	32
ABCDEFGHIJKLMNOPQRSTUVWXYZ	26
× 0	2
0123456789	10

Мощность алфавита (N) – полное количество символов алфавита.



Все символы несут одинаковую информацию:

информационная
емкость символа

$$I = \log_2 N$$

мощность
алфавита

Алфавитный подход

Задача. Определить объем информации в сообщении

ПРИВЕТВАСЯ

для кодирования которого используется русский алфавит (только заглавные буквы).



Решение:

- считаем все символы (здесь **10** символов)
- мощность алфавита – **32 символа** ($32=2^5$)
- 1 символ несет **5 бит** информации

Ответ: $10 \cdot 5 \text{ бит} = 50 \text{ бит}$

Количество информации в тексте

А теперь для того, чтобы найти количество информации во всем тексте, **нужно посчитать число символов в нем и умножить на I.**

Посчитаем количество информации на одной странице книги.



Пусть страница содержит **50 строк**. В каждой строке — **60 символов**. Значит, на странице помещается **$50 \times 60 = 3000$** знаков.

Тогда объем информации будет равен:
 $5,755 \times 3000 = 17265$ бит.

При алфавитном подходе к измерению информации количество информации зависит не от содержания, а от размера текста и мощности алфавита.



Задача:

Определите информационный объем страницы книги, если для записи текста использовались только заглавные буквы русского алфавита, кроме буквы Ё.

Решение:

1. $N = 32$
2. $2^I = N$
3. $2^I = 32$
4. $I = 5$
5. На странице 3000 знаков, тогда объем информации = $3000 * 5 = 15000$ бит.

Достаточный алфавит

Удобнее всего измерять информацию, когда размер алфавита N равен целой степени двойки.

Например, если $N=16$, то каждый символ несет 4 бита информации потому, что $2^4 = 16$. А если $N=32$, то один символ «весит» 5 бит.

Ограничения на максимальный размер алфавита теоретически не существует. Однако есть алфавит, который можно назвать **достаточным**. С ним мы скоро встретимся при работе с компьютером. Это алфавит **мощностью 256 символов**.

В алфавит такого размера можно поместить все практически необходимые символы: латинские и русские буквы, цифры, знаки арифметических операций, всевозможные скобки, знаки препинания....

Поскольку $256 = 2^8$, то один символ этого алфавита «весит» 8 бит. Причем 8 бит информации — это настолько характерная величина, что ей даже присвоили свое название — байт.

1 байт = 8 бит

Количество информации в тексте

Сегодня очень многие люди для подготовки писем, документов, статей, книг и пр. используют компьютерные текстовые редакторы. Компьютерные редакторы, в основном, работают с алфавитом размером 256 символов.

В этом случае легко подсчитать объем информации в тексте. Если 1 символ алфавита несет 1 байт информации, то надо просто сосчитать количество символов; полученное число даст информационный объем текста в байтах.



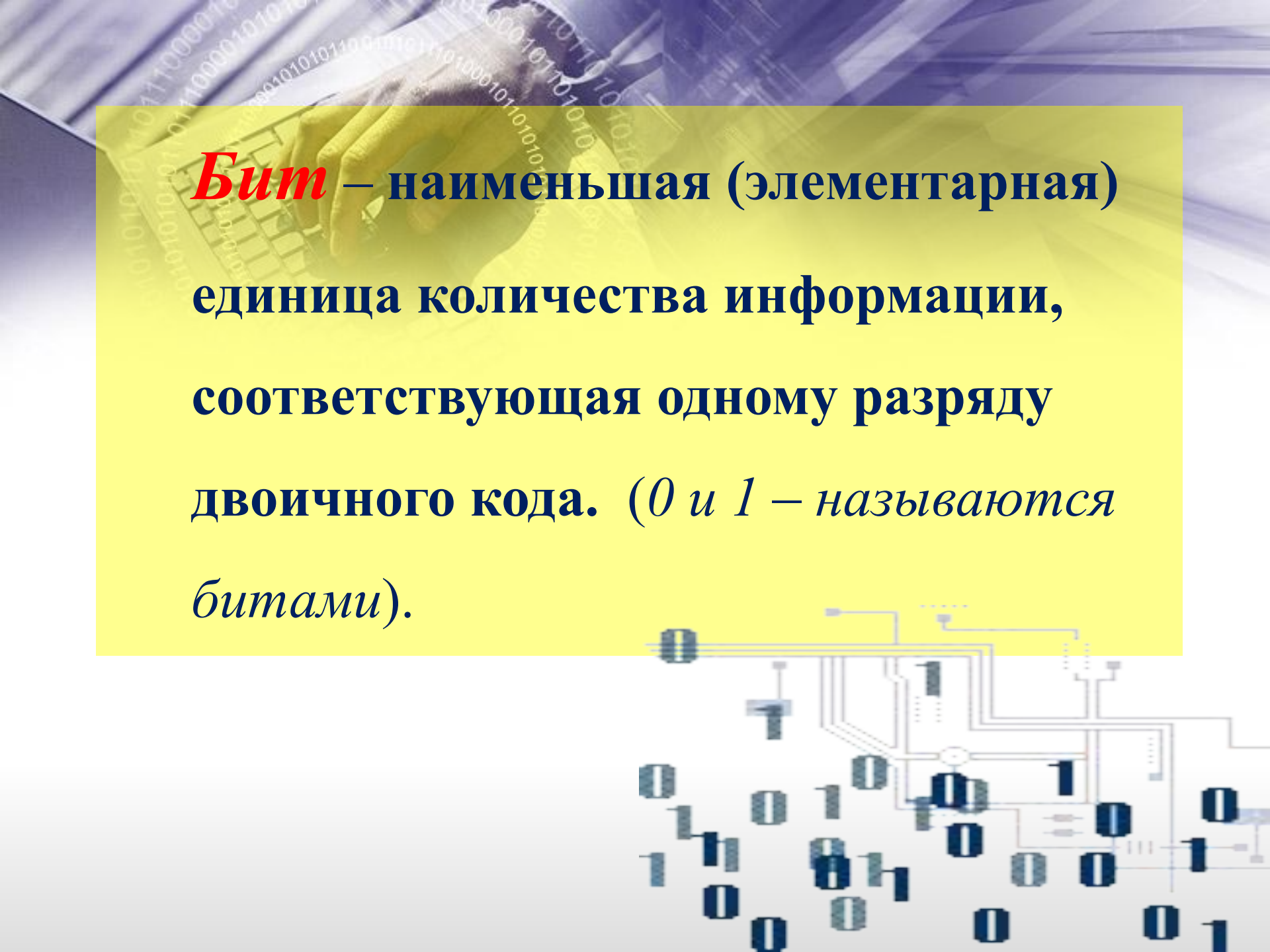
Пусть небольшая книжка, сделанная с помощью компьютера, содержит *150 страниц*; на каждой странице — *40 строк*, в каждой строке — *60 символов*.

Значит страница содержит $40 \times 60 = 2400$ **байт информации**.

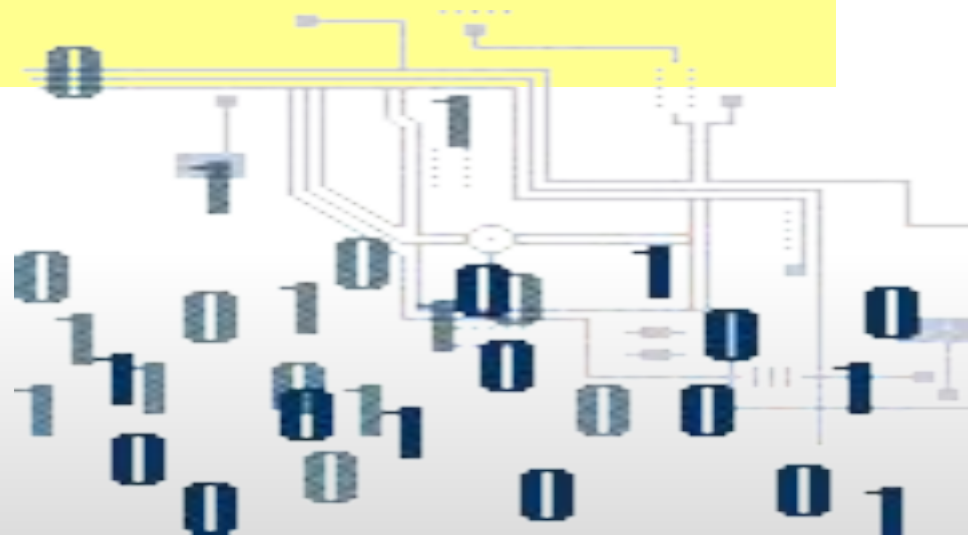
Объем всей информации в книге: $2400 \times 150 = 360\,000$ **байт**.



Единицы измерения информация



Бит – наименьшая (элементарная) единица количества информации, соответствующая одному разряду двоичного кода. (*0 и 1 – называются битами*).





Группа из 8 бит называется ***байтом***
(byte – binary term – двоичный элемент)

Байт – основная единица
измерения информации,
занесенная в систему СИ

Измерения в байтах

Десятичная приставка

Двоичная приставка

Название	Символ	Степень	Название	Символ		Степень
				МЭК	ГОСТ	
килобайт	KB	10^3	кибибайт	KiB	Килобит	2^{10}
мегабайт	MB	10^6	мебибайт	MiB	Мегабит	2^{20}
гигабайт	GB	10^9	гибибайт	GiB	Гигабит	2^{30}
терабайт	TB	10^{12}	тебибайт	TiB	Терабит	2^{40}
петабайт	PB	10^{15}	пебибайт	PiB		2^{50}
эксабайт	EB	10^{18}	эксбибайт	EiB		2^{60}
зеттабайт	ZB	10^{21}	зебибайт	ZiB		2^{70}
йоттабайт	YB	10^{24}	йобибайт	YiB		2^{80}



The end

