

Дискретное (цифровое)
представление текстовой,
графической, звуковой информации
и видеоинформации

Кодирование и декодирование

Для обмена информацией с другими людьми человек использует естественные языки. Наряду с *естественными* языками были разработаны *формальные* языки для профессионального применения их в какой-либо сфере. Представление информации с помощью какого-либо языка часто называют кодированием.

Код — набор символов (условных обозначений) для представления информации.

Код — система условных знаков (символов) для передачи, обработки и хранения информации (сообщения).

Кодирование — процесс представления информации (сообщения) в виде кода.

Все множество символов, используемых для кодирования, называется *алфавитом кодирования*. Например, в памяти компьютера любая информация кодируется с помощью двоичного алфавита, содержащего всего два символа: **0** и **1**.

Способы кодирования информации

Для кодирования одной и той же информации могут быть использованы разные способы; их выбор зависит от ряда обстоятельств: цели кодирования, условий, имеющихся средств.

Если надо записать текст в темпе речи — используем стенографию; если надо передать текст за границу — используем английский алфавит; если надо представить текст в виде, понятном для грамотного русского человека, — записываем его по правилам грамматики русского языка.

«Добрый день, Дима!»

«Dobryi den, Dima»

Двоичное кодирование в компьютере

Вся информация, которую обрабатывает компьютер должна быть представлена двоичным кодом с помощью двух цифр: **0** и **1**. *Эти два символа принято называть двоичными цифрами или битами.*

С помощью двух цифр 0 и 1 можно закодировать любое сообщение. Это явилось причиной того, что в компьютере обязательно должно быть организовано два важных процесса: кодирование и декодирование.

Кодирование – преобразование входной информации в форму, воспринимаемую компьютером, т.е. двоичный код.

Почему двоичное кодирование

С точки зрения технической реализации использование двоичной системы счисления для кодирования информации оказалось намного более простым, чем применение других способов. Действительно, удобно кодировать информацию в виде последовательности нулей и единиц, если представить эти значения как два возможных устойчивых состояния электронного элемента:

0 – отсутствие электрического сигнала;

1 – наличие электрического сигнала.

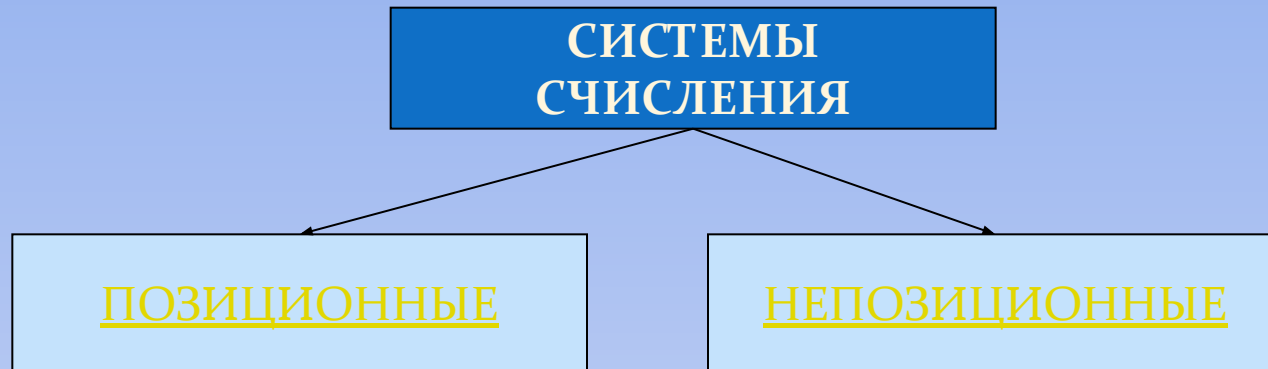
Способы кодирования и декодирования информации в компьютере, в первую очередь, зависит от вида информации, а именно, что должно кодироваться: числа, текст, графические изображения или звук.

Система счисления

Для записи информации о количестве объектов используются числа. Числа записываются с помощью набора специальных символов.

Система счисления — способ записи чисел с помощью набора специальных знаков, называемых цифрами.

Виды систем счисления



В позиционных системах счисления величина, обозначаемая цифрой в записи числа, зависит от её положения в числе (позиции).

В непозиционных системах счисления величина, которую обозначает цифра, не зависит от положения в числе.

Непозиционные системы счисления

Каноническим примером фактически *непозиционной системы* счисления является *римская*, в которой в качестве цифр используются латинские буквы:

I обозначает 1, V - 5, X - 10, L - 50, C - 100, D - 500, M -1000.

Натуральные числа записываются при помощи повторения этих цифр.

Например, II = 1 + 1 = 2, здесь символ I обозначает 1 независимо от места в числе.

Для правильной записи больших чисел римскими цифрами необходимо сначала записать число тысяч, затем сотен, затем десятков и, наконец, единиц.

Пример: число 2988. Две тысячи MM, девять сотен CM, восемьдесят LXXX, восемь VIII. Запишем их вместе: MCMLXXXVIII.

$MMCMLXXXVIII = 1000+1000+(1000-100)+(50+10+10+10)+5+1+1+1 = 2988$

Для изображения чисел в непозиционной системе счисления нельзя ограничиться конечным набором цифр. Кроме того, выполнение арифметических действий в них крайне неудобно.

Древнеегипетская десятичная непозиционная система счисления.

- Примерно в третьем тысячелетии до нашей эры древние египтяне придумали свою числовую систему, в которой для обозначения ключевых чисел 1, 10, 100 и т.д. использовались специальные значки — иероглифы.
- Все остальные числа составлялись из этих ключевых при помощи операции сложения. Система счисления Древнего Египта является десятичной, но непозиционной.

Алфавитные системы счисления.

- Более совершенными непозиционными системами счисления были алфавитные системы. К числу таких систем счисления относились греческая, славянская, финикийская и другие. В них числа от 1 до 9, целые количества десятков (от 10 до 90) и целые количества сотен (от 100 до 900) обозначались буквами алфавита.
- В алфавитной системе счисления Древней Греции числа 1, 2, ..., 9 обозначались первыми девятью буквами греческого алфавита, например $\alpha = 1$, $\beta = 2$, $\gamma = 3$ и т.д. Для обозначения чисел 10, 20, ..., 90 применялись следующие 9 букв ($\iota = 10$, $\kappa = 20$, $\lambda = 30$, $\mu = 40$ и т.д.), а для обозначения чисел 100, 200, ..., 900 — последние 9 букв ($\rho = 100$, $\sigma = 200$, $\tau = 300$ и т.д.). Например, число 141 обозначалось $\rho\mu\alpha$.
- У славянских народов числовые значения букв установились в порядке славянского алфавита, который использовал сначала глаголицу, а затем кириллицу. Подробнее с происхождением и развитием русской письменности можно ознакомиться на сайте

Древнерусская алфавитная система счисления

БУКВЫ КИРИЛЛИЦЫ	ЦИФРОВОЕ ЗНАЧЕНИЕ КИРИЛЛИЦЫ	БУКВЫ ГЛАГОЛИЦЫ	ЦИФРОВОЕ ЗНАЧЕНИЕ ГЛАГОЛИЦЫ	КИРИЛЛИЧЕСКОЕ НАЗВАНИЕ
А	1	Ѧ	1	Азь
Б		Ѣ	2	Буки
В	2	ѦѢ	3	Вѣди
Г	3	ѦѢѦ	4	Глаголь
Д	4	ѦѢѢ	5	Добро
Е	5	ѦѢѢѢ	6	Есть
Ж		ѦѢѢѢѢ	7	Живѣте
З	6	ѦѢѢѢѢѢ	8	Зѣло
З	7	ѦѢѢѢѢѢѢ	9	Земля
Н	8	ѦѢѢѢѢѢѢѢ	10	Иже
І	10	ѦѢѢѢѢѢѢѢѢ	20	І

Позиционные системы счисления

В позиционных системах счисления величина, обозначаемая цифрой в записи числа, зависит от её положения в числе (позиции).

Количество используемых цифр называется основанием системы счисления.

Например, 11 – это одиннадцать, а не два: $1 + 1 = 2$ (сравните с римской системой счисления). Здесь символ 1 имеет различное значение в зависимости от позиции в числе.

Первые позиционные системы счисления

Самой первой такой системой, когда счетным "прибором" служили пальцы рук, была *пятеричная*.



Некоторые племена на филиппинских островах используют ее и в наши дни, а в цивилизованных странах ее реликт, как считают специалисты, сохранился только в виде школьной пятибалльной шкалы оценок.

Двенадцатеричная система счисления

Следующей после пятеричной возникла *двенадцатеричная* система счисления. Возникла она в древнем Шумере. Некоторые учёные полагают, что такая система возникала у них из подсчёта фаланг на руке большим пальцем.

Широкое распространение получила двенадцатеричная система счисления в XIX веке. На ее широкое использование в прошлом явно указывают названия числительных во многих языках, а также сохранившиеся в ряде стран способы отсчета времени, денег и соотношения между некоторыми единицами измерения. Год состоит из 12 месяцев, а половина суток состоит из 12 часов.

Элементом двенадцатеричной системы в современности может служить счёт дюжинами. Первые три степени числа 12 имеют собственные названия: 1 дюжина = 12 штук; 1 гросс = 12 дюжин = 144 штуки; 1 масса = 12 гроссов = 144 дюжины = 1728 штук.

Английский фунт состоит из 12 шиллингов.

Шестидесятеричная система счисления

Следующая позиционная система счисления была придумана еще в Древнем Вавилоне, причем вавилонская нумерация была *шестидесятеричная*, т.е. в ней использовалось шестьдесят цифр!

В более позднее время использовалась арабами, а также древними и средневековыми астрономами. Шестидесятеричная система счисления, как считают исследователи, является синтезом уже вышеупомянутых пятеричной и двенадцатеричной систем.

Какие позиционные системы счисления используются сейчас?

В настоящее время наиболее распространены десятичная, двоичная, восьмеричная и шестнадцатеричная системы счисления.

Двоичная, восьмеричная (в настоящее время вытесняется шестнадцатеричной) и шестнадцатеричная система часто используется в областях, связанных с цифровыми устройствами, программировании и вообще компьютерной документации.

Современные компьютерные системы оперируют информацией представленной в цифровой форме.

Десятичная система счисления

Десятичная система счисления

— позиционная система
счисления по основанию 10.

Предполагается, что основание
10 связано с количеством
пальцев рук у человека.

Наиболее распространённая
система счисления в мире.

Для записи чисел используются
символы 0, 1, 2, 3, 4, 5, 6, 7, 8, 9,
называемые арабскими
цифрами.

Современные цифры	Арабские цифры	Индийские цифры
0	٠	०
1	١	१
2	٢	२
3	٣	३
4	٤	४
5	٥	५
6	٦	६
7	٧	७
8	٨	८
9	٩	९

Двоичная система счисления

Двоичная система счисления — позиционная система счисления с основанием 2.

Используются цифры 0 и 1.

Двоичная система используется в цифровых устройствах, поскольку является наиболее простой и удовлетворяет требованиям:

- Чем меньше значений существует в системе, тем проще изготовить отдельные элементы.
- Чем меньше количество состояний у элемента, тем выше помехоустойчивость и тем быстрее он может работать.
- Простота создания таблиц сложения и умножения — основных действий над числами

Алфавит десятичной, двоичной, восьмеричной и шестнадцатеричной систем счисления

Система счисления	Основание	Алфавит цифр
Десятичная	10	0, 1, 2, 3, 4, 5, 6, 7, 8, 9
Двоичная	2	0, 1
Восьмеричная	8	0, 1, 2, 3, 4, 5, 6, 7
Шестнадцатеричная	16	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F

Соответствие десятичной, двоичной, восьмеричной и шестнадцатеричной систем счисления

p=10	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
p=2	0	1	10	11	100	101	110	111	1000	1001	1010	1011	1100	1101	1110	1111	10000
p=8	0	1	2	3	4	5	6	7	10	11	12	13	14	15	16	17	20
p=16	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	10

Количество используемых цифр называется основанием системы счисления.

При одновременной работе с несколькими системами счисления для их различения основание системы обычно указывается в виде нижнего индекса, который записывается в десятичной системе:

123_{10} — это число 123 в десятичной системе счисления;

11011_2 — то же число, но в двоичной системе.

Двоичное число 11011 можно расписать в виде: $11011_2 = 1 \cdot 2^4 + 1 \cdot 2^3 + 0 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0$.

Перевод чисел из одной системы счисления в другую

Перевод из десятичной системы счисления в систему счисления с основанием p осуществляется последовательным делением десятичного числа и его десятичных частных на p , а затем выписыванием последнего частного и остатков в обратном порядке.

Переведем десятичное число 99_{10} в двоичную систему счисления (основание системы счисления $p=2$). В итоге получили $1000011_2 = 99_{10}$

$$\begin{array}{r} 99 \mid 2 \\ \hline 49 \mid 2 \\ \hline 1 \quad 48 \mid 24 \mid 2 \\ \hline 1 \quad 24 \mid 12 \mid 2 \\ \hline 0 \quad 12 \mid 6 \mid 2 \\ \hline 0 \quad 6 \mid 3 \mid 2 \\ \hline 0 \quad 3 \mid 1 \\ \hline 0 \quad 2 \mid 1 \\ \hline 0 \end{array}$$

Числа в компьютере

Числа в компьютере хранятся и обрабатываются в двоичной системе счисления.

Последовательность нулей и единиц называют двоичным кодом.

Специфической особенности представления чисел в памяти компьютера рассмотрим на других уроках по теме «**системы счисления**».

Кодирование текстовой информации

- В традиционных кодировках для кодирования одного символа используется 8 бит. Легко подсчитать по формуле 2.3, что такой 8-разрядный код позволяет закодировать 256 различных символов.
- Присвоение символу определенного числового кода – это вопрос соглашения. В качестве международного стандарта принята кодовая таблица ASCII (American Standard Code for Information Interchange), кодирующая первую половину символов с числовыми кодами от 0 до 127 (коды от 0 до 32 отведены не символам, а функциональным клавишам).
- Национальные стандарты кодировочных таблиц включают международную часть кодовой таблицы без изменений, а во второй половине содержат коды национальных алфавитов, символы псевдографики и некоторые математические знаки. К сожалению, в настоящее время существуют пять различных кодировок кириллицы (КОИ8-Р, Windows.MS-DOS, Macintosh и ISO), что вызывает дополнительные трудности при работе с русскоязычными документами.
- Хронологически одним из первых стандартов кодирования русских букв на компьютерах был КОИ8 ("Код обмена информацией, 8-битный"). Эта кодировка применялась еще в 70-ые годы на компьютерах серии ЕС ЭВМ, а с середины 80-х стала использоваться в первых русифицированных версиях операционной системы UNIX.
- Наиболее распространенной в настоящее время является кодировка Microsoft Windows, обозначаемая сокращением CP1251 ("CP" означает "Code Page", "кодовая страница").

Кодирование текстовой информации

- От начала 90-ых годов, времени господства операционной системы MS DOS, остается кодировка CP866. Компьютеры фирмы Apple, работающие под управлением операционной системы Mac OS, используют свою собственную кодировку Mac. Кроме того, Международная организация по стандартизации (International Standards Organization, ISO) утвердила в качестве стандарта для русского языка еще одну кодировку под названием ISO 8859-5.

- В конце 90-ых годов появился новый международный стандарт Unicode, который отводит под один символ не один байт, а два, и поэтому с его помощью можно закодировать не 256, а 65536 различных символов. Полная спецификация стандарта Unicode включает в себя все существующие, вымершие и искусственно созданные алфавиты мира, а также множество математических, музыкальных, химических и прочих СИМВОЛОВ.

- **Пример** . Представьте в форме шестнадцатеричного кода слово «ЭВМ» во всех пяти кодировках. Воспользуйтесь CD-ROM для получения кодировочных таблиц CP866, Mac и ISO и компьютерным калькулятором для перевода чисел из десятичной в шестнадцатеричную систему

Кодирование текстовой информации

- Последовательности десятичных кодов слова «ЭВМ» в различных кодировках составляем на основе кодировочных таблиц:
- КОИ8-Р: 252 247 237
- CP1251: 221 194 204
- CP866: 157 130 140
- Mac: 157 130 140
- ISO: 205 178 188
- Переводим с помощью калькулятора последовательности кодов из десятичной системы в шестнадцатеричную:
- КОИ8-Р: FC F7 ED
- CP1251: DD C2 CC
- CP866: 9D 82 8C
- Mac: 9D 82 8C
- ISO: CD B2 BC
- Для преобразования русскоязычных текстовых документов из одной кодировки в другую используются специальные программы-конверторы. Одной из таких программ является текстовый редактор Hieroglyph, который позволяет осуществлять перевод набранного текста из одной кодировки в другую и даже использовать различные кодировки в одном тексте.

Международная кодировка ASCII

sp 32	! 33	" 34	# 35	\$ 36	% 37	& 38	' 39	(40) 41	* 42	+ 43	, 44	- 45	. 46	/ 47
0 48	1 49	2 50	3 51	4 52	5 53	6 54	7 55	8 56	9 57	: 58	; 59	< 60	= 61	> 62	? 63
@ 64	A 65	B 66	C 67	D 68	E 69	F 70	G 71	H 72	I 73	J 74	K 75	L 76	M 77	N 78	O 79
P 80	Q 81	R 82	S 83	T 84	U 85	V 86	W 87	X 88	Y 89	Z 90	[91	\ 92] 93	^ 94	_ 95
` 96	a 97	b 98	c 99	d 100	e 101	f 102	g 103	h 104	i 105	j 106	k 107	l 108	m 109	n 110	o 111
p 112	q 113	r 114	s 115	t 116	u 117	v 118	w 119	x 120	y 121	z 122	{ 123	 124	} 125	~ 126	

Кодировка КОИ8-Р

—		Г	Г	└	└	┌	┌	┐	┐	┘	■	■	■	■	■
128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
▨	▨	▨	┌	■	●	√	≈	≤	≥	nbsp	┘	◦	2	•	÷
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
=		F	ё	П	Р	Т	П	П	Е	Ц	Ц	┘	┘	┘	┘
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
┘	┘	┘	Ё	┘	┘	┘	┘	┘	┘	┘	┘	┘	┘	┘	©
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
Ю	а	б	ц	д	е	ф	г	х	и	й	к	л	м	н	о
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
п	я	р	с	т	у	ж	в	ь	ы	з	ш	э	щ	ч	ъ
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
Ю	А	Б	Ц	Д	Е	Ф	Г	Х	И	Й	К	Л	М	Н	О
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
П	Я	Р	С	Т	У	Ж	В	Ь	Ы	З	Ш	Э	Щ	Ч	Ъ
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

Кодировка CP1251

Á	à	,	è	„	…	†	‡	€	‰	É	<	й	Й	Ó	Ú
128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
á	‘	’	“	”	•	–	—	€	™	é	>	ò	й	ó	ú
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
nbsp	ÿ	Ы	Э	И	Ы	!	§	€	©	Ю	«	¬	shy	®	Я
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
°	±	Ы	Э	’	µ	¶	•	ё	№	Ю	»	э	Ю	Я	Я
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

Представление видеоинформации

В последнее время компьютер все чаще используется для работы с видеоинформацией. Простейшей такой работой является просмотр кинофильмов и видеоклипов. Следует четко представлять, что обработка видеоинформации требует очень высокого быстродействия компьютерной системы.

Что представляет собой *фильм* с точки зрения информатики? Прежде всего, это *сочетание звуковой и графической информации*. Кроме того, для создания на экране эффекта движения используется дискретная по своей сути технология быстрой смены статических картинок. Исследования показали, что если за одну секунду сменяется более 10-12 кадров, то человеческий глаз воспринимает изменения на них как непрерывные.

Представление видеоинформации

Казалось бы, если проблемы кодирования статической графики и звука решены, то сохранить видеоизображение уже не составит труда.

Но это только на первый взгляд, поскольку при использовании традиционных методов сохранения информации электронная версия фильма получится слишком большой.

Достаточно очевидное усовершенствование состоит в том, чтобы первый кадр запомнить целиком (в литературе его принято называть ключевым), а в следующих сохранять лишь отличия от начального кадра (разностные кадры).

Некоторые форматы видеофайлов

Существует множество различных форматов представления видеоданных.

- В среде Windows, например, уже более 10 лет применяется формат Video for Windows, базирующийся на универсальных файлах с расширением AVI (Audio Video Interleave – чередование аудио и видео).
- Более универсальным является мультимедийный формат Quick Time, первоначально возникший на компьютерах Apple.
- Все большее распространение в последнее время получают системы сжатия видеоизображений, допускающие некоторые незаметные для глаза искажения изображения с целью повышения степени сжатия. Наиболее известным стандартом подобного класса служит MPEG (Motion Picture Expert Group). Методы, применяемые в MPEG, непросты для понимания и опираются на достаточно сложную математику.
- Больше распространение получила технология под названием DivX (Digital Video Express). Благодаря DivX удалось достигнуть степени сжатия, позволившей вместить качественную запись полнометражного фильма на один компакт-диск – сжать 4,7 Гб DVD-фильма до 650 Мб.

Мультимедиа

Мультимедиа (multimedia, от англ. multi - много и media - носитель, среда) - совокупность компьютерных технологий, одновременно использующих несколько информационных сред: текст, графику, видео, фотографию, анимацию, звуковые эффекты, высококачественное звуковое сопровождение.

Под словом «мультимедиа» понимают воздействие на пользователя по нескольким информационным каналам одновременно. Можно еще сказать так: **мультимедиа – это объединение изображения на экране компьютера (в том числе и графической анимации и видеокадров) с текстом и звуковым сопровождением.**

Наибольшее распространение системы мультимедиа получили в области обучения, рекламы, развлечений.

Задания:

1. Запишите число 1945 в римской системе счисления.
2. Рассчитайте необходимый объем видеопамати для графического режима: разрешение экрана 800x600, качество цветопередачи 16 бит.
3. Используя таблицу символов, записать последовательность десятичных числовых кодов в кодировке Windows для своих ФИО, названия улицы, по которой проживаете. Таблица символов отображается в редакторе MS Word с помощью команды: вкладка Вставка→Символ→Другие символы.

В поле Шрифт выбираете Times New Roman, в поле из выбираете кириллица. Например, для буквы «А» (русской заглавной) код знака– 192.

Пример:

И	В	А	Н	О	В		А	Р	Т	Е	М
200	194	192	205	206	194		192	208	210	197	204
	П	Е	Т	Р	О	В	И	Ч			
	207	197	210	208	206	194	200	215			