

Лекция
«Кодирование
информации»

Специальность «ЛД»

1. Основные понятия кодирования

- **Информация** — это сведения об объектах и явлениях окружающего мира, уменьшающие степень неопределённости знаний об этих объектах или явлениях.
- **Источник информации** — это объект, порождающий информацию и представляющий её в виде сообщения.
- **Приемник информации** — это объект, принимающий сообщение и способный правильно его интерпретировать.

Под словом «**кодирование**» понимают процесс представления информации, удобный для её хранения и/или передачи.

- **Код** — система условных знаков (символов), предназначенных для представления информации в соответствии с определенными правилами.
- **Кодирование** — переход от одной формы представления информации к другой, наиболее удобной для её хранения, передачи или обработки.
- **Декодирование** — процесс по восстановлению первоначальной формы представления информации, т. е. операция, обратная кодированию.



Наиболее распространенные цели кодирования:

- 1) **экономность** (сократить запись);
- 2) **надежность** (засекретить информацию);
- 3) **удобство обработки или восприятия.**

Существуют три основных способа кодирования текста:

- 1) **графический** — с помощью специальных рисунков или значков;
- 2) **числовой** — с помощью чисел;
- 3) **символьный** — с помощью символов того же алфавита, что и исходный текст.

«Здравствуй, дорогой Саша!»

«Zdravstvui, dorogoi Sasha!»

Из-за кодирования символов «вЕС»
первого сообщения> второго

«тридцать пять умножить на сто
двадцать семь»

или «35x127»

2. Основные понятия

шифрования

- **Шифрование** — метод защиты любой информации от несанкционированного доступа, просмотра, а также её использования, основанный на преобразовании данных в зашифрованный формат.
- **Дешифрование** — процесс обратного преобразования, при котором восстанавливается исходный текст
- **Криптография** — это наука о методах и принципах передачи и приема зашифрованной с помощью специальных ключей информации.

3. Представление текстовой информации

Все символы кодируются числами, и текст представляется в виде набора чисел — *кодов символов*.

При выводе текста на экран монитора или принтер необходимо восстановить изображения всех символов, составляющих данный текст. Для этого используются так называемые *кодовые (кодировочные) таблицы*, в которых каждому коду ставится в соответствие изображение символа.

В программировании наиболее часто используются *однобайтовые кодировки*, в которых код каждого символа занимает ровно 1 байт (или 8 бит). Общее количество различных символов составляет 256

3.1. Кодировка ASCII

Наиболее распространенной системой байтового кодирования является система (таблица) ASCII (**American Standard Code for Information Interchange** — американский стандартный код для обмена информацией), созданная в 1963 году.

В таблице ASCII всего 256 позиций. Каждому двоичному коду от 0 до 255 в таблице ставится в соответствие один символ. Сама таблица делится на две части. Первая часть с 0-го по 127-й символ называется *основной* таблицей ASCII и является неизменной для всех стран. Вторая часть – до 255-го символа – называется *расширенной* таблицей ASCII, она уникальна для каждой страны, так как в этой части хранятся символы, специфичные для этой страны, например буквы национального алфавита.

0 -	16 - ▶	32 -	48 - 0	64 - @	80 - P	96 - '	112 - p
1 - ☺	17 - ◀	33 - !	49 - 1	65 - A	81 - Q	97 - a	113 - q
2 - ☹	18 - ↕	34 - "	50 - 2	66 - B	82 - R	98 - b	114 - r
3 - ♥	19 - !!	35 - #	51 - 3	67 - C	83 - S	99 - c	115 - s
4 - ♦	20 - ♣	36 - \$	52 - 4	68 - D	84 - T	100 - d	116 - t
5 - ♣	21 - ♠	37 - %	53 - 5	69 - E	85 - U	101 - e	117 - u
6 - ♠	22 - —	38 - &	54 - 6	70 - F	86 - V	102 - f	118 - v
7 -	23 - ↕	39 - '	55 - 7	71 - G	87 - W	103 - g	119 - w
8 -	24 - ↑	40 - (56 - 8	72 - H	88 - X	104 - h	120 - x
9 -	25 - ↓	41 -)	57 - 9	73 - I	89 - Y	105 - i	121 - y
10 -	26 - →	42 - *	58 - :	74 - J	90 - Z	106 - j	122 - z
11 -	27 - ←	43 - +	59 - ;	75 - K	91 - [107 - k	123 - {
12 -	28 - └	44 - ,	60 - <	76 - L	92 - \	108 - l	124 -
13 -	29 - ↔	45 - -	61 - =	77 - M	93 - j	109 - m	125 - }
14 - 🎵	30 - ▲	46 - .	62 - >	78 - N	94 - ^	110 - n	126 - ~
15 - ⚙	31 - ▼	47 - /	63 - ?	79 - O	95 - ÷	111 - o	127 - ⏏
16 - ▶	32 -	48 - 0	64 - @	80 - P	96 -	112 - p	

3.2 Кодировки КОИ-8 и Windows-1251

Для русского языка наиболее распространенными являются однобайтовые кодовые таблицы CP-866, Windows-1251 и КОИ-8. В них первые 128 символов совпадают с ASCII-кодировкой, а русские буквы размещены во второй части таблицы, однако коды русских букв в этих кодировках различны.

Несовпадение кодовых таблиц приводит к тому, что текст, набранный в одной кодировке, будет нечитабельным в другой.

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

3.3. Кодировка UNICODE

К концу 1980-х годов стандартом стали 8-битные коды, при этом существовало множество разных 8-битных кодировок и постоянно появлялись все новые.

В результате появилась необходимость решения нескольких задач:

- Проблема «кракозябр» (отображения документов в неправильной кодировке).
- Проблема ограниченности набора символов.
- Проблема преобразования одной кодировки в другую.

Было признано необходимым создание единой «широкой» кодировки фиксированной ширины. Использование 32-битных кодов казалось слишком расточительным, поэтому было решено использовать 16-битные коды.

Первая версия «Юникода» представляла собой кодировку с фиксированным размером кода в 16 бит, общее число кодов в ней было 2^{16} (65 536).

Полная 32-битная таблица «Юникод» включает символы практически всех современных письменностей.

С академическими целями в эту кодировочную таблицу добавлены многие исторические письменности, в том числе: руны, древнегреческая письменность, египетские иероглифы, клинопись, письменность майя, этрусский алфавит. В «Юникоде» представлен широкий набор математических и музыкальных символов, а также пиктограмм.

4. Представление графической информации

Создавать и хранить графические объекты в компьютере можно двумя способами: как **растровое** или как **векторное** изображение. Для каждого типа изображения используется свой способ кодирования.



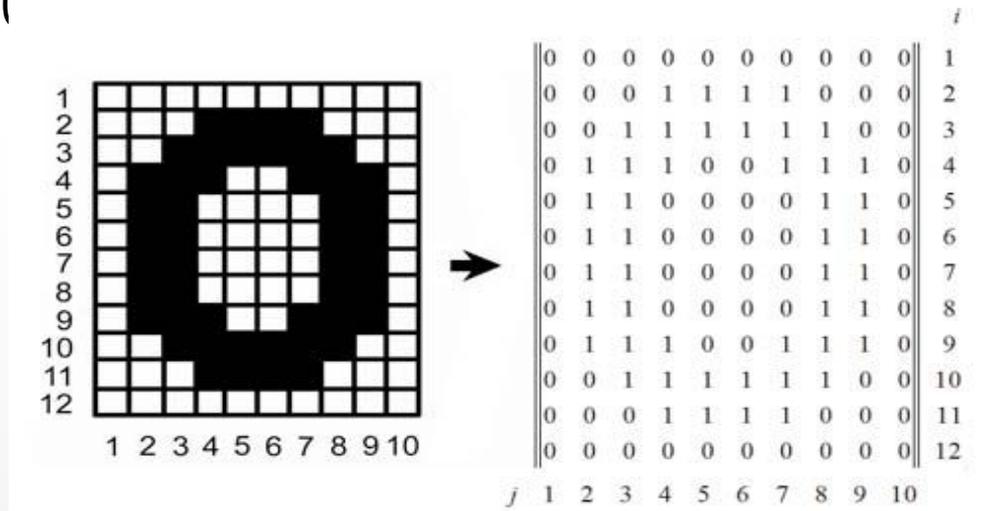
РАСТР
.jpeg .gif .png



ВЕКТОР
.svg

Растровое изображение - совокупность точек, используемых для его отображения на экране монитора.

Объём растрового изображения определяется как произведение количества точек и информационного объёма одной точки, который зависит от количества возможных цветов. Для черно-белого изображения информационный объём одной точки равен 1 биту, так как точка может быть либо чёрной, либо белой, что можно закодировать одной из двух цифр — 0 и 1.



Векторное изображение представляет собой совокупность графических примитивов. Каждый примитив состоит из элементарных отрезков кривых, параметры которых (координаты узловых точек, радиус кривизны и пр.) описываются математическими формулами.

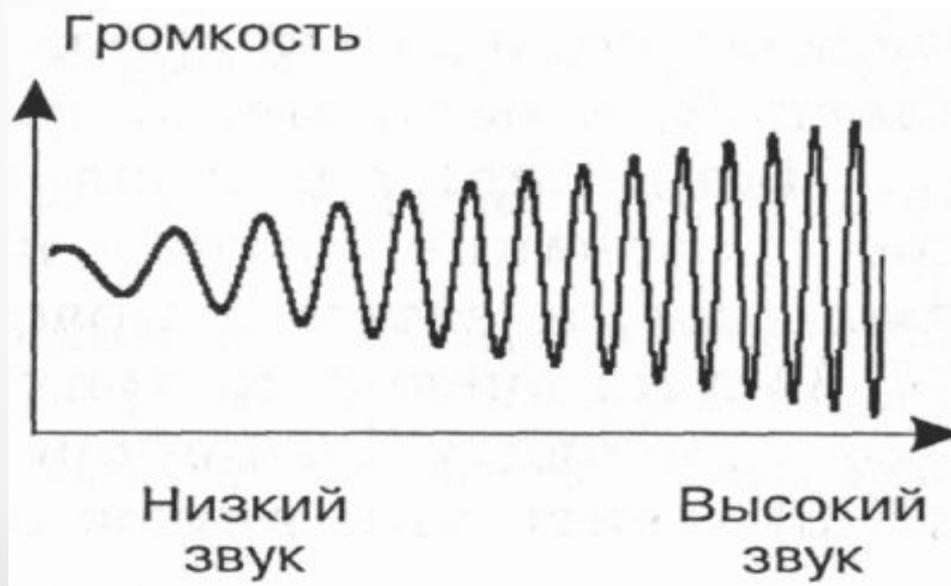
Информация о векторном рисунке кодируется обычным способом, как хранятся тексты, формулы, числа, т. е. хранится не графическое изображение, а только координаты и характеристики изображения его деталей. Поэтому для хранения векторных изображений требуется существенно меньше памяти, чем растровых изображений.

Цвет на Web-страницах кодируется в виде **RGB**-кода в шестнадцатеричной системе: **#RRGGBB**, где **RR**, **GG** и **BB** — яркости красного, зеленого и синего, записанные в виде двух шестнадцатеричных цифр; это позволяет закодировать 256 значений от 0 (0016) до 255 (FF16) для каждой составляющей.

Color Name	Color Code	Color Name	Color Code
Red	#FF0000	White	#FFFFFF
Cyan	#00FFFF	Silver	#C0C0C0
Blue	#0000FF	Grey	#808080
DarkBlue	#0000A0	Black	#000000
LightBlue	#ADD8E6	Orange	#FFA500
Purple	#800080	Brown	#A52A2A
Yellow	#FFFF00	Maroon	#800000
Lime	#00FF00	Green	#008000
Fuchsia	#FF00FF	Olive	#808000

5. Представление звуковой информации

- Звук представляет собой непрерывный сигнал — звуковую волну с меняющейся амплитудой и частотой.
- Чем больше амплитуда сигнала, тем он громче для человека.



Количество бит, отводимое на один звуковой сигнал, называют глубиной кодирования звука.

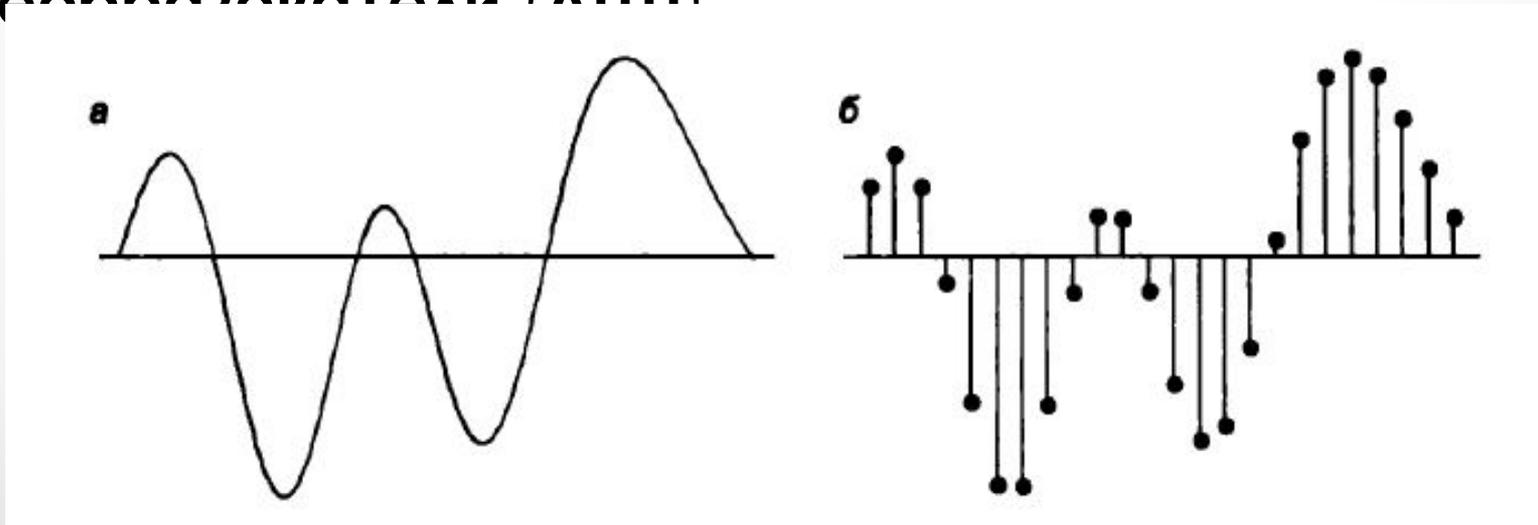
Современные звуковые карты обеспечивают 16-, 32- или 64-битную глубину кодирования звука.

При кодировании звуковой информации непрерывный сигнал заменяется **дискретным**, то есть превращается в последовательность электрических импульсов (двоичных нулей и единиц).

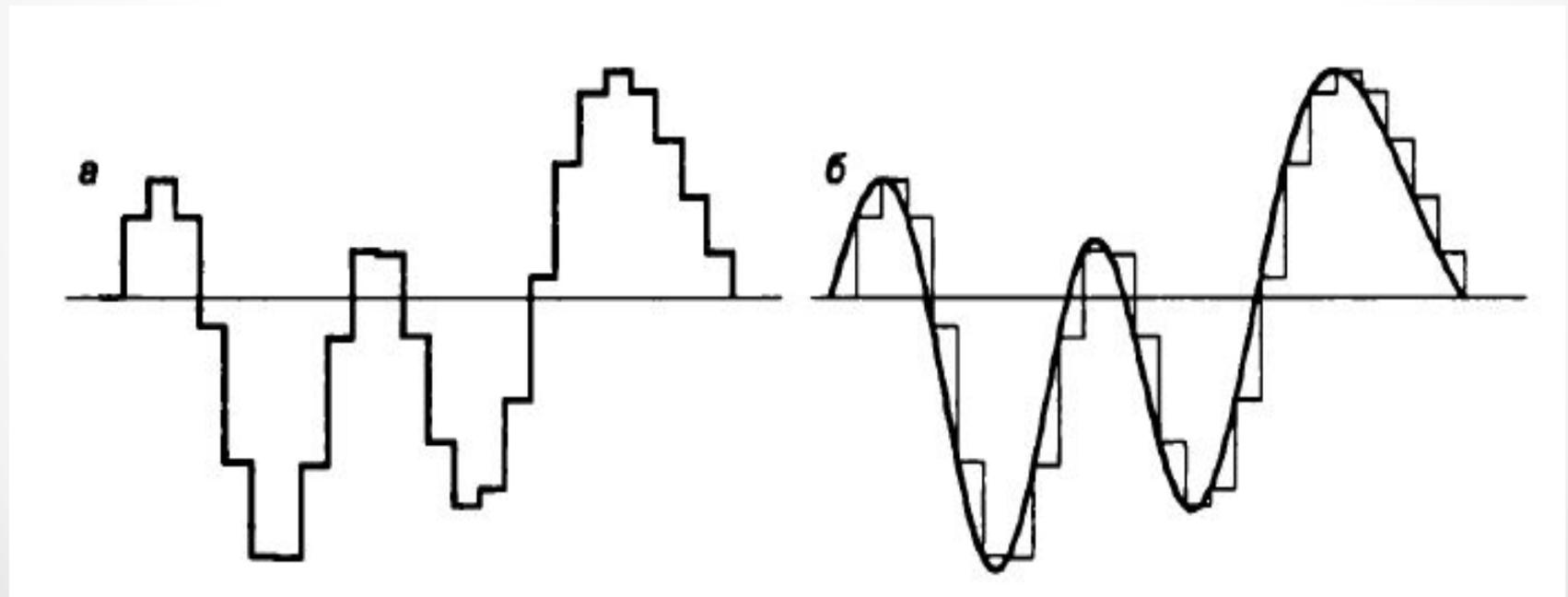
Процесс перевода звуковых сигналов от непрерывной формы представления к дискретной, цифровой форме называют оцифровкой.

Существуют различные методы кодирования звуковой информации двоичным кодом, среди которых можно выделить два основных направления: метод **FM** и метод **Wave-Table**.

Метод FM (*Frequency Modulation*) основан на том, что теоретически любой сложный звук можно разложить на последовательность простейших гармонических сигналов разных частот, каждый из которых представляет собой правильную синусоиду, и, следовательно, может быть описан кодом. Разложение звуковых сигналов в гармонические ряды и представление в виде дискретных цифровых сигналов выполняют специальные устройства — **аналогово-цифровые преобразователи (АЦП)**



Обратное преобразование для воспроизведения звука, закодированного числовым кодом, выполняют **цифро-аналоговые преобразователи (ЦАП)**. Процесс преобразования звука представлен на рис. ниже. Данный метод кодирования не даёт хорошего качества звучания, но обеспечивает компактный код.



Таблично-волновой метод (*Wave-Table*) основан на том, что в заранее подготовленных таблицах хранятся образцы звуков окружающего мира, музыкальных инструментов и т. Д, называемые **СЭМПЛАМИ**.

Числовые коды выражают высоту тона, продолжительность и интенсивность звука и прочие параметры, характеризующие особенности звука. Поскольку в качестве образцов используются «реальные» звуки, качество звука, полученного в результате синтеза, получается очень высоким и приближается к качеству звучания реальных музыкальных инструментов.

Звуковые файлы имеют несколько форматов. Наиболее популярные из них **MIDI, WAV, MP3**.





