

Кодирование текстовой информации



Кодирование текстов:

- Текст в памяти компьютера хранится как последовательность двоичных кодов символов. Изображения символов хранятся в отдельных шрифтовых файлах (расширения: **fon**, **ttf**, **otf**).
- Для предоставления текста в памяти компьютера необходимо представить его в виде двоичного код, т.е. в виде последовательности из **0** и **1**



Кодирование

- это процесс преобразования информации в форму, воспринимаемую компьютером, т.е. двоичный код.

Декодирование

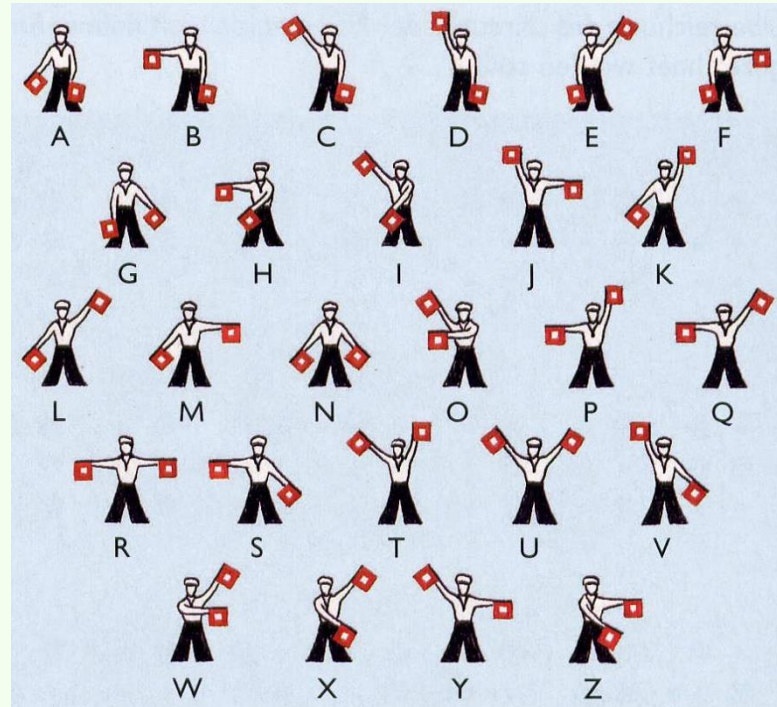
- это процесс обратный кодированию, который заключается в том, что происходит преобразование данных из двоичного кода в форму, понятную человеку.



Сигнальные коды

A	. -	S	. . .
B	- . . .	T	-
C	- . - .	U	. . -
D	- . .	V	. . . -
E	.	W	. - -
F	. . - .	X	- . . -
G	- - .	Y	- . - -
H		Z	- - . .
I	. .	1	. - - - -
J	. - - -	2	. . - - -
K	- . -	3	. . . - -
L	. - . .	4	 -
M	- -	5	
N	- .	6	-
O	- - -	7	- - . . .
P	. - - .	8	- - - . .
Q	- - . -	9	- - - - .
R	. - .	0	- - - - -

Азбука Морзе (1837 г)



Флажковая азбука

Способы кодирования

1) Замена буквы её порядковым номером в алфавите:

- $A \rightarrow 1, Б \rightarrow 2, В \rightarrow 3, \dots, Я \rightarrow 33$

Замена буквы следующей за ней по алфавиту:

- $A \rightarrow B, B \rightarrow B, B \rightarrow \Gamma, \dots, Я \rightarrow A$

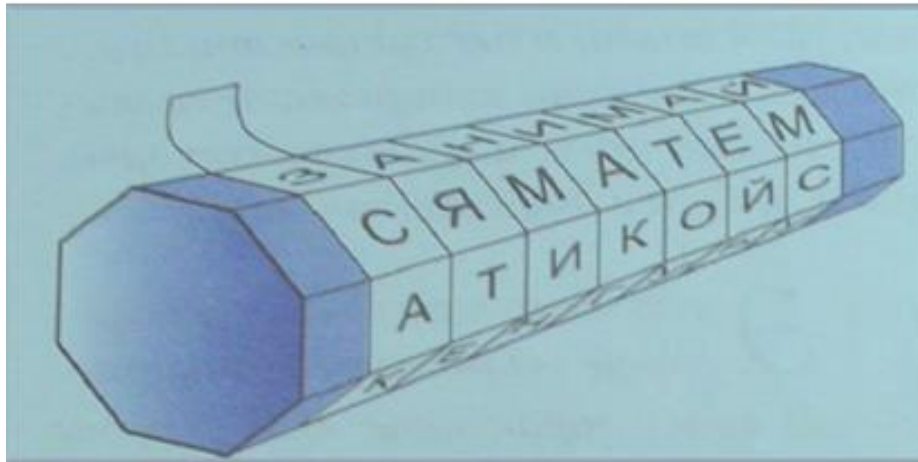
Замена буквы на букву, номер которой на 2 больше исходного:

- $A \rightarrow B, B \rightarrow \Gamma, B \rightarrow D, \dots, Я \rightarrow B$



Шифр Сцитала

- Для прочтения шифровки нужно не только знать систему засекречивания, но и обладать ключом в виде палочки, принятого диаметра. Зная тип шифра, но, не имея ключа, расшифровать сообщение было сложно. Этот шифр именовался сцитала по названию стержня, на который наматывались свитки папируса, что указывает на его происхождение. Он был весьма популярен в Спарте и много раз совершенствовался в позднейшие времена



Шифр Энея

- **«диск Энея»** - небольшой диск, с просверленными вдоль его края отверстиями соответствующими буквам алфавита. Для того, чтобы зашифровать текст «шишка», нить вдевали в отверстия так, чтобы по ходу нити последовательность отверстий была такой же, как у соответствующих букв в тексте.
- Получатель сообщения, вытягивая эту нить, мог прочитать все буквы, правда, в обратном порядке.



Шифр Цезаря

- Шифр Цезаря предельно прост: в нем каждая буква сообщения заменялась на следующую за ней по алфавиту с указанием сдвига

А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П	Р	С	Т	...	Ь	Ъ	Э	Ю	Я
Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П	Р	С	Т	У	Ф	Х		Ю	Я	А	Б	В

3



Шифр Вижинера

- При шифровании исходного сообщения его выписывают в строку, а под ним записывают ключевое слово. Если ключ оказался короче сообщения, то его циклически повторяют. В процессе шифрования очередная буква шифр-текста находится на пересечении столбца, определяемого шифруемой буквой, и строки, определяемой значением ключа.

• Сообщение	П Р И Л Е Т А Ю	С Е Д Ь М О Г О
• Ключ	А М Б Р О З И Я	А М Б Р О З И Я
• Шифр-текст	П Ъ Й Ы У Щ И Э	С С Е К Ъ Х Л Н



Таблица кодировки

Таблица, в которой всем символам компьютерного алфавита поставлены в соответствие порядковые номера (коды)

Международным стандартом для ПК стала таблица ASCII (Американский стандартный код для информационного обмена).

32 пробел	48 0	64 @	80 P	96 `	112 p
33 !	49 1	65 A	81 Q	97 a	113 q
34 "	50 2	66 B	82 R	98 b	114 r
35 #	51 3	67 C	83 S	99 c	115 s
36 \$	52 4	68 D	84 T	100 d	116 t
37 %	53 5	69 E	85 U	101 e	117 u
38 &	54 6	70 F	86 V	102 f	118 v
39 '	55 7	71 G	87 W	103 g	119 w
40 (	56 8	72 H	88 X	104 h	120 x
41)	57 9	73 I	89 Y	105 i	121 y
42 *	58 :	74 J	90 Z	106 j	122 z
43 +	59 ;	75 K	91 [	107 k	123 {
44 ,	60 <	76 L	92 \	108 l	124
45 -	61 =	77 M	93]	109 m	125 }
46 .	62 >	78 N	94 ^	110 n	126 ~
47 /	63 ?	79 O	95 _	111 o	127

Первую половину кодовой страницы (коды от **0** до **127**) занимает **стандартная** таблица ASCII, а **вторую** (коды от **128** до **256**) — **символы национальных алфавитов** (например, русские буквы).

Для русского языка существуют несколько кодовых страниц, которые были разработаны для разных операционных систем

в системе Windows

[illegible]

Windows-1251 (CP-1251)

в Unix-совместимых
операционных
системах и
электронной почте

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

KOI8-R

в системе MS DOS

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
8	A 125	Б 129	В 130	Г 131	Д 132	Е 133	Ж 134	З 135	И 136	Й 137	К 138	Л 139	М 140	Н 141	О 142	П 143
9	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
A	а 160	б 161	в 162	г 163	д 164	е 165	ж 166	з 167	и 168	й 169	к 170	л 171	м 172	н 173	о 174	п 175
B	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
C	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
D	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
E	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
F	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

CP-866

на компьютерах
фирмы Apple
(Макинтош и др.)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
В.	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
Р.	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
А.	†	°	€	£	§	•	І	Ж	©	™	Ъ	Ы	Ь	Э	Ю	Я
В.	00	≤	≥	і	μ	đ	Ј	Є	І	І	Љ	Љ	Љ	Љ	Љ	Љ
С.	j	S	~	√	f	≈	А	α	α	...	Ъ	Ы	К	К	К	К
В.	—	α	α	α	α	α	α	α	α	α	Ц	ц	№	№	№	№
Е.	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
Р.	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я

MacCyrillic

Windows-1251	KOI8-R
Привет, Вася!	oПХБЕР, бЮЯЪ!
рТЙЧЕФ, чБУС!	Привет, Вася!

Стандарт UNICODE

Для того чтобы в одном документе можно было использовать более 256 символов, в 1991 году был принят стандарт кодирования символов UNICODE, который включает знаки любых существующих (и даже некоторых мёртвых) языков, математические и музыкальные символы и др.

Windows

- **UTF-16**
- В ней все наиболее важные символы кодируются с помощью 16 бит (2 байт), а редко используемые — с помощью 4 байт.

Unix - Linux

- **UTF-8**
- В ней все символы, входящие в таблицу ASCII, кодируются с помощью 1 байта, а другие символы могут занимать от 2 до 4 байт. Если значительную часть текста составляют латинские буквы и цифры, такой подход позволяет значительно уменьшить объём файла по сравнению с UTF-16.



Задание:

- Выписать в тетрадь основные понятия
- Выполнить задания:

1. Какое из предложенных слов может быть закодировано следующим образом:

- a) секрет
- b) сеттер
- c) сервер
- d) турнир



2. Слово **КУРСОР** закодировано следующим образом: #*!^+!

Как будет закодировано слово **УРОК** этим кодом?

3. Для шифровки каждой буквы слова используются двузначные числа. Известно, что буква «к» закодирована числом 15 . Среди слов **торт**, **ёжик**, **станок**, **беседа** есть слова, кодируемые последовательностями цифр: 35291815, 303113241115. Какая последовательность цифр является кодом слова «**китёнок**» ?

Задание:

4. Вася шифрует русские слова, записывая вместо каждой буквы её номер в алфавите (без пробелов). Номера букв даны в таблице.

Определите, какое сообщение закодировано в строчке **191014031613**

А 01	Й 11	У 21	Э 31
Б 02	К 12	Ф 22	Ю 32
В 03	Л 13	Х 23	Я 33
Г 04	М 14	Ц 24	
Д 05	Н 15	Ч 25	
Е 06	О 16	Ш 26	
Ё 07	П 17	Щ 27	
Ж 08	Р 18	Ъ 28	
З 09	С 19	Ы 29	
И 10	Т 20	Ь 30	

5. От разведчика была получена следующая шифрованная радиোগрамма, переданная с использованием азбуки Морзе:

• • - • • • - • - - • - • • • • -

При передаче радиограммы было потеряно разбиение на буквы, но известно, что в радиограмме использовались только следующие буквы:

Т	А	У	Ж	Х
-	• -	• • -	• • • -	• • • •

Запишите в ответе расшифрованную радиограмму.