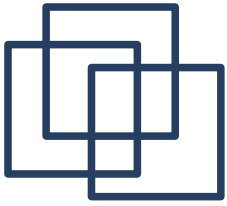
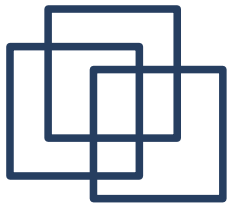


Кодирование СИМВОЛОВ

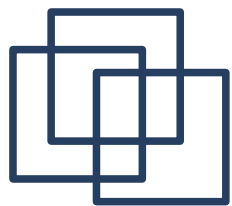


Основу любого языка составляет **алфавит** - конечный набор различных знаков (символов) любой природы, из которых складывается сообщение.



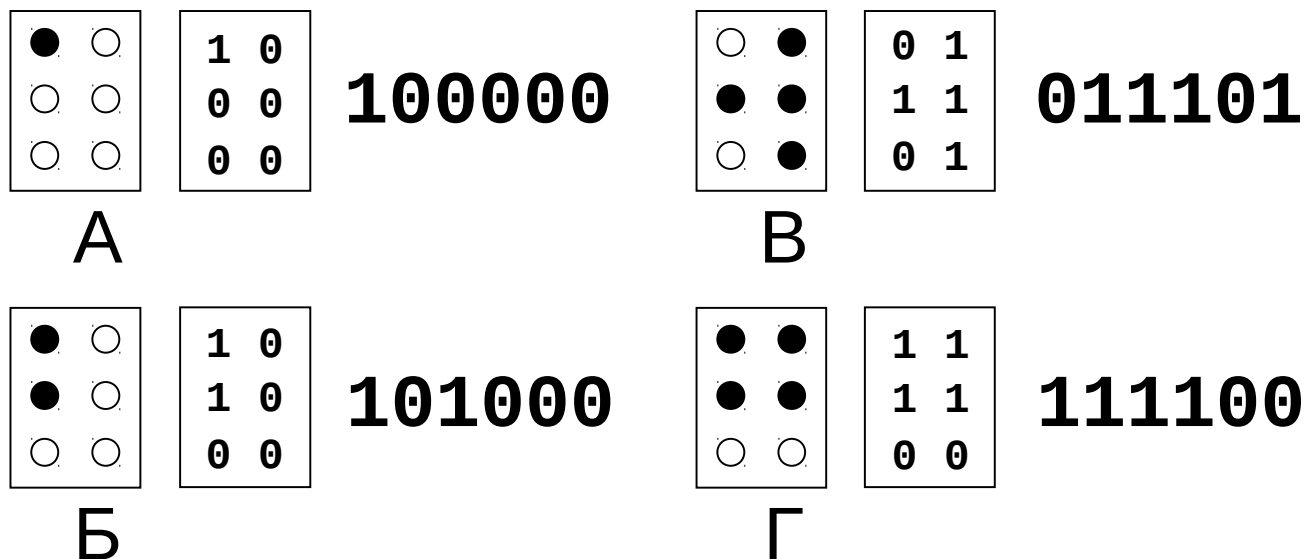
Компьютер может обрабатывать только информацию, представленную в числовой форме. Вся другая информация (звуки, изображения, показания приборов и т.д.) для обработки на компьютере должна быть преобразована в числовую форму.

Под термином «**кодирование**» понимают переход от одной формы представления информации к другой, более удобной для хранения, передачи или обработки.



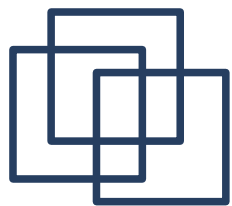
Кодирование символов

Система Брайля:



Общий подход:

- Определить мощность алфавита **N**
- Выделить необходимое **число битов k на символ**: **$2^k \geq N$**
- Сопоставить каждому символу **код** – число от 0 до **$2^k - 1$**
- Перевести коды **в двоичную систему**



Кодирование символов

Текстовый файл

- на экране (символы)
- в памяти – коды



1000001_2

65

1000010_2

66

1000011_2

67

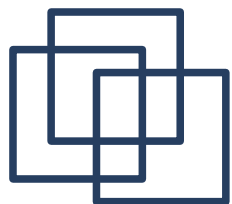
1000100_2

68



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

Файлы со шрифтами: ***.fon**, ***.ttf**, ***.otf**



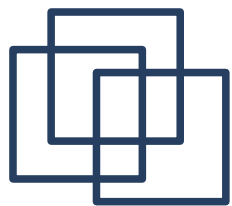
Основная

кодированная таблица ASCII:

Для кодирования текстовой информации принят международный стандарт **ASCII** (American Standard Code for Information Interchange), в кодовой таблице которого первоначально было зарезервировано 128 (от 0 до 127) 7-ми разрядных кодов для кодирования:

$$2^k = N; 2^7 = 128$$

При этом первые 33 кода (от 0 до 32) являются управляющими символами. Другие коды основной таблицы ASCII используются для кодирования букв латинского алфавита, цифр, знаков препинания и математических символов.

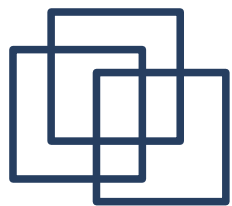


Расширения

кодовой таблицы ASCII:

Существующие 8-битные расширения основной таблицы ASCII в дополнение к первым 128 символам позволяют кодировать не только символы псевдографики или математические символы, но и буквы национальных алфавитов. Для этого используются десятичные коды от 128 до 255.

$$2^k = N; 2^8 = 256$$



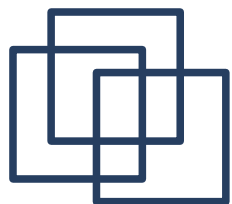
Проблема

кодировки кириллицы:

При построении расширений основной таблицы ASCII важно использовать согласованный стандарт, чтобы избежать несовместимости при представлении символов.

К сожалению, так не произошло при кодировании букв русского (кириллического) алфавита. Возникло несколько вариантов расширения основной таблицы ASCII с русскими (кириллическими) буквами.

Ситуация осложнилась тем, что в СССР (а затем в России, на Украине, в Белоруссии и т.д.) вовремя не был принят единый стандарт подобного расширения и несколько расширенных кодовых таблиц стали стандартом де-факто.

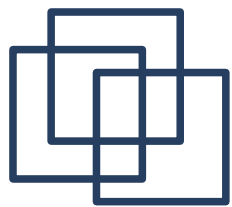


История создания

кодových таблиц для кириллицы:

В 70-х г. в СССР одними из первых были разработаны и утвержден в качестве государственного стандарта (ГОСТ 19768—74) 8-битный код обмена и обработки информации (КОИ-8) — кодовая таблица **CP-KOI8**, которая вводилась в действие с 01.01.1975 до 01.01.1980.

К тому времени относится также использование нескольких (заслуженно забытых сегодня) внутриведомственных кодовых таблиц.

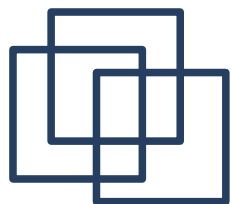


История создания

КОДОВЫХ

В конце 80-х годов с введением периферийных устройств предложена та кодировка, исполняющая локализации управления о принятая в качестве одн из стандартов СР-

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0		⊙	⊖	♥	♦	♣	♠	●		○						
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	▸	◀					—		↑	↓	→	←		↔	▲	▼
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
2		!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
5	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95
6	'	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111
7	p	q	r	s	t	u	v	w	x	y	z	{		}	~	▯
	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127
8	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
9	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
A	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	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	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
F	Ё	ё	≥	≤		┐	┌	└	┘	┐	┌	└	┘	┐	┌	└
	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

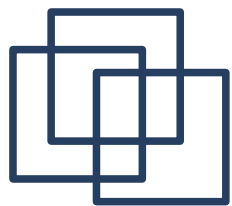


История создания

кодových таблиц для кириллицы:

В 1988 г. Международной организацией стандартов был утвержден стандарт **ISO-8859-5** кодового представления кириллицы, используемыми UNIX-платформы.

Одновременно с широким использованием операционной системы UNIX и развитием российских компьютерных сетей происходило распространение еще одной кодовой таблицы **CP KOI-8R**, базирующейся на ГОСТ 19768—74

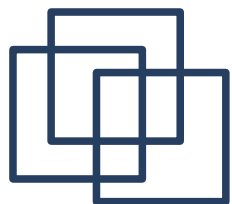


История создания

кодových таблиц для кириллицы:

Появление и
интерфейса
привело к ма
кодовой таб
базируется р
программного
корпорацией М

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0				©	Ё	§	Є	.		°						
1		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
3		!	"	#	\$	%	&	(	)	^	+					/
4	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
5	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
6	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
7	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
8	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
9	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
A	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95
B	'	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
C	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111
D	p	q	r	s	t	u	v	w	x	y	z	{		}	~	□
E	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127
F	Ђ	Ѓ	Ѕ	Ї	Љ	Њ	Ћ	Ќ		‰	Љ	«	Њ	Ћ	Ќ	Љ
	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
	ђ	‘	’	”	”		—	—		™	љ	»	њ	ќ	ћ	џ
	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
A		Ў	ў	Ј	Ѡ	Ѓ	Ѕ	Ї	Љ	Њ	Ћ	«		-	®	Ї
B	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
C	°	±	І	і	Ѓ	μ		.	ё	№	є	»	ј	Ѕ	ѕ	ї
D	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
E	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
F	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



История создания

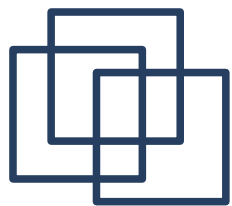
кодových таблиц для кириллицы:

Последняя (по времени) кодовая таблица была разработана в связи с выходом на российский компьютерной рынок компании Apple, которая для локализации рабочих станций Macintosh, работающих под операционной системой MacOS, ввела свою собственную кодовую таблицу **CP10007 (X-Mac-Cyrillic)**.



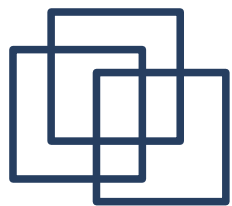
Пример «кракозябры»





8-битные кодировки

- 1 байт на символ – файлы небольшого размера!
 - просто обрабатывать в программах
- нельзя использовать символы разных кодовых страниц одновременно (русские и французские буквы, и т.п.)
 - неясно, в какой кодировке текст (перебор вариантов!)
 - для каждой кодировки нужен свой шрифт (изображения символов)



Стандарт UNICODE

1 112 064 знаков, используются около **100 000**

Windows: **UTF-16**

16 битов на распространённые символы,
32 бита на редко встречающиеся

Linux: **UTF-8**

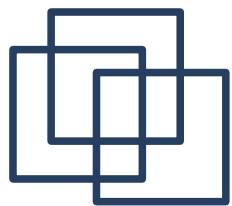
8 битов на символ для ASCII,
от 16 до 48 бита на остальные



- совместимость с ASCII
- более экономична, чем UTF-16, если много символов ASCII

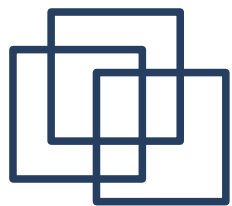


2010 г. – 50% сайтов использовали UTF-8!



Задание на дом

§ 15



Задание 1.

В таблице ниже представлена часть кодовой таблицы ASCII:

Символ	1	5	J	K	P	j	k
Десятичный код	49	53	74	75	80	106	107
Шестнадцатеричный код	31	35	4A	4B	50	6A	6B

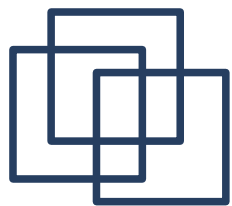
Каков шестнадцатеричный код символа «p» ?

1) 70_{16}

2) 85_{16}

3) $6F_{16}$

4) CE_{16}



Задание 2.

В таблице ниже представлена часть кодовой таблицы ASCII:

Символ	С	Т	У	Я	с	т	у
Десятичный код	145	146	147	159	225	226	227
Шестнадцатеричный код	91	92	93	9F	E1	E2	E3

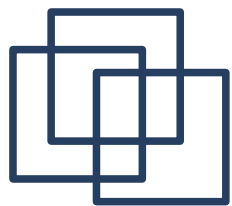
Каков шестнадцатеричный код символа «я» ?

1) $2A_{16}$

2) DF_{16}

3) EF_{16}

4) 180_{16}



Задание 3.

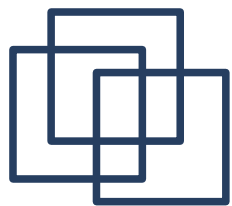
Автоматическое устройство осуществило перекодировку информационного сообщения на русском языке, первоначально записанного в 16-битном коде *Unicode*, в 8-битную кодировку *КОИ-8*. При этом информационное сообщение уменьшилось на 160 бит. Какова длина сообщения в символах?

1) 10

2) 20

3) 40

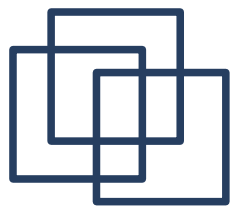
4) 160



Задание 4.

Автоматическое устройство осуществило перекодировку информационного сообщения на русском языке длиной 8 символов, первоначально записанного в 16-битном коде *Unicode*, в 8-битную кодировку *КОИ-8*. При этом информационный объем сообщения уменьшился на

- 1) 8 байт
- 2) 8 бит
- 3) 64 байта
- 4) 128 бит



Задание 5.

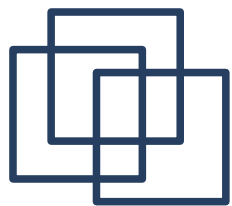
Автоматическое устройство осуществило перекодировку информационного сообщения на русском языке, первоначально записанного в 16-битном коде *Unicode*, в 8-битную кодировку *КОИ-8*. При этом информационное сообщение уменьшилось на 240 бит. Какова длина сообщения в символах?

1) 15

2) 30

3) 60

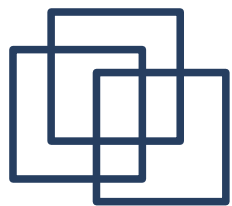
4) 240



Задание 6.

Автоматическое устройство осуществило перекодировку информационного сообщения на русском языке длиной в 20 символов, первоначально записанного в 16-битном коде *Unicode*, в 8-битную кодировку *КОИ-8*. При этом информационное сообщение уменьшилось на

- 1) 320 бит
- 2) 20 бит
- 3) 160 байт
- 4) 20 байт



Задание 7.

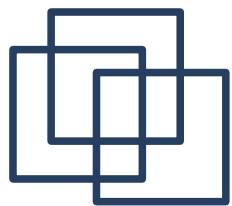
Информационный объем сообщения равен 40960 бит.
Чему равен объем этого сообщения в Кбайтах?

1) 5

2) 8

3) 32

4) 12



Задание 8.

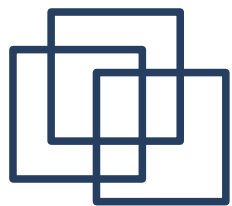
Информационный объем сообщения равен 40960 бит.
Чему равен объем этого сообщения в Кбайтах?

1) 5

2) 8

3) 32

4) 12



Задание 9.

В таблице ниже представлена часть кодовой таблицы ASCII:

Символ	1	5	A	B	a	b	r
Десятичный код	49	53	65	66	97	98	114
Шестнадцатеричный код	31	35	41	42	61	62	72

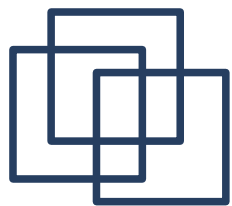
Каков шестнадцатеричный код символа «R» ?

1) $A0_{16}$

2) 72_{16}

3) $A2_{16}$

4) 52_{16}

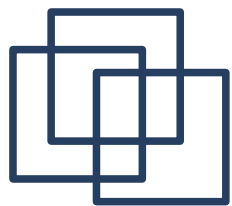


Задание 10.

В кодировке КОИ-8 каждый символ кодируется одним байтом. Определите информационный объем следующего сообщения в этой кодировке:

У сильного всегда бессильный виноват.

- 1) 33 байта
- 2) 37 бит
- 3) 256 бит
- 4) 296 бит



Задание 11.

В таблице ниже представлена часть кодовой таблицы ASCII:

Символ	1	3	A	Y	Z	y	z
Десятичный код	49	51	65	89	90	121	122
Восьмеричный код	61	63	101	131	132	171	172

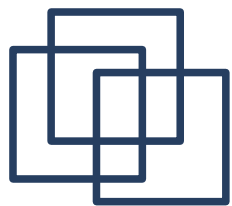
Каков восьмеричный код символа «а» ?

1) 97_8

2) 65_8

3) 101_8

4) 141_8



Задание 12.

В таблице ниже представлена часть кодовой таблицы ASCII:

Символ	1	3	X	Y	Z	c	z
Десятичный код	49	51	88	89	90	99	122
Восьмеричный код	61	63	130	131	132	143	172

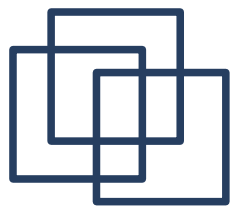
Каков восьмеричный код символа «C» ?

1) 67_8

2) 103_8

3) 97_8

4) 143_8



Задание 13.

1) Текстовый документ, состоящий из 5120 символов, хранился в 8-битной кодировке КОИ-8. Этот документ был преобразован в 32-битную кодировку. Укажите, какое дополнительное количество Кбайт потребуется для хранения документа. В ответе запишите только число.

15

2) Текстовый документ хранился в 8-битной кодировке КОИ-8. Этот документ был преобразован в 16-битную кодировку Unicode, при этом размер памяти, необходимой для хранения документа увеличился на 4 Кбайт. При этом хранится только последовательность кодов символов. Укажите, сколько символов в документе. В ответе запишите только число.

4069

3) Автоматическое устройство осуществило перекодировку информационного сообщения на русском языке, первоначально записанного в 2-байтном коде Unicode, в 8-битную кодировку КОИ-8. При этом длина сообщения уменьшилась на 600 бит. Укажите, сколько символов было в сообщении.

75