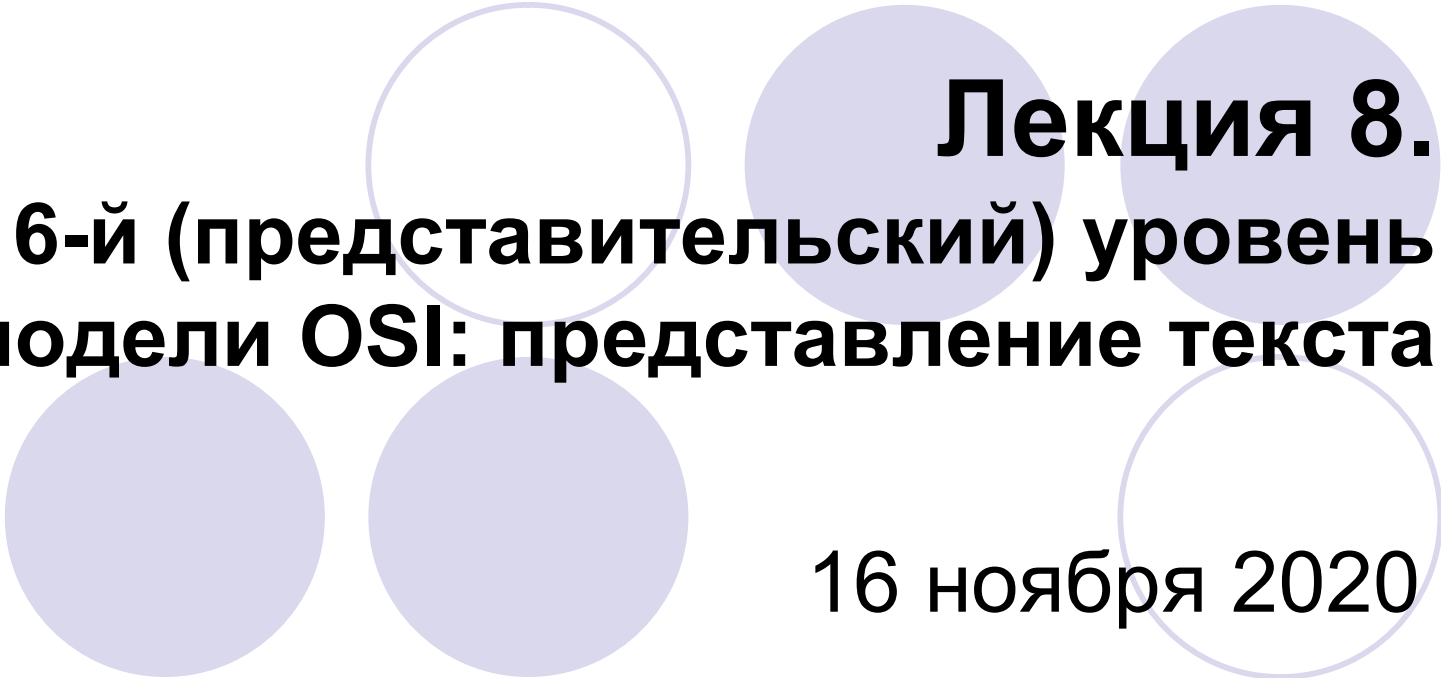


КПИ им. Игоря Сикорского, каф. микроэлектроники.

М. Р. Домбругов. Информатика-1.

Персональные компьютеры и основы сетевых технологий



Лекция 8.

6-й (представительский) уровень модели OSI: представление текста

16 ноября 2020

7-уровневая модель OSI

Тип данных	Уровень (layer)	Функции
Данные	7. Прикладной (application)	Доступ к сетевым службам
	6. Представительский (presentation)	Представление и шифрование данных
	5. Сеансовый (session)	Управление сеансом связи
Сегменты	4. Транспортный (transport)	Прямая связь между конечными пунктами и надежность
Пакеты	3. Сетевой (network)	Определение маршрута и логическая адресация
Кадры	2. Канальный (data link)	Физическая адресация
Биты	1. Физический (physical)	Работа со средой передачи, сигналами и двоичными данными

6-й уровень модели OSI

Представление и шифрование данных

- **Представление текста**
- - " - чисел
- - " - звука
- - " - изображений
- - " - видео
- Оптимальное кодирование и архиваторы
- Шифрование данных
- ...

- 1-байтовая кодировка
- 8-битная кодировка кириллицы
- Unicode
- RTF – Rich Text Format
- Компьютерные шрифты и их характеристики
- Векторные и растровые шрифты
- Пропорциональные и моноширинные шрифты
- Стилистические группы шрифтов

1-байтовая кодировка

ASCII (American Standard Code for Information Interchange)

Введен в действие институтом стандартизации США (ANSI – American National Standard Institute) в 1963 г.

7-битная кодировка, коды символов $00_{16} \dots 7F_{16}$ (0 ... 127)

ASCII Code Chart

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
1	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
2		!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL

1-байтовая кодировка

Для управления принтером используются первые 32 кода:

- BS (08)** Backspace (Возврат на один символ). Вернуть печатающую каретку назад на одну позицию.
- HT (09)** Horizontal Tabulation (Горизонтальное Табулирование). Продвинуть каретку до следующей предписанной 'позиции табуляции'.
- LF (0A)** Line Feed (Перевод строки). Провернуть бумагоопорный вал на одну строку вниз.
- VT (0B)** Vertical Tabulation (Вертикальное Табулирование). Провернуть вал к следующей группе строк.
- FF (0C)** Form Feed (Перевод страницы). Провернуть вал к исходной позиции следующей страницы.
- CR (0D)** Carriage Return (Возврат каретки). Вернуть каретку к исходной (крайней левой) позиции текущей строки.

1-байтовая кодировка

Пример неформатированного текста (plain-text):

Jingle bells, jingle bells
Jingle all the way
Oh, what fun it is to ride
In a one horse open sleigh

управляющие коды

00000000:	4A 69 6E 67 6C 65 20 62	65 6C 6C 73 2C 20 6A 69	Jingle bells, ji
00000010:	6E 67 6C 65 20 62 65 6C	6C 73 0D 0A 4A 69 6E 67	ngle bells..Jing
00000020:	6C 65 20 61 6C 6C 20 74	68 65 20 77 61 79 0D 0A	le all the way..
00000030:	4F 68 2C 20 77 68 61 74	20 66 75 6E 20 69 74 20	Oh, what fun it
00000040:	69 73 20 74 6F 20 72 69	64 65 0D 0A 49 6E 20 61	is to ride..In a
00000050:	20 6F 6E 65 20 68 6F 72	73 65 20 6F 70 65 6E 20	one horse open
00000060:	73 6C 65 69 67 68		sleigh

1-байтовая кодировка

Русификация ASCII:

коды **КОИ-7**
(код обмена
информацией,
7 бит).

Наборы символов
КОИ-7Н1 и
КОИ-7Н2

(Существовал еще
КОИ-7Н0,
совпадавший во
всем с ASCII,
но с заменой \$ на ¤)

20	21	22	23	24	25	26	27	28	29	2A	2B	2C	2D	2E	2F
	!	"	#	¤	%	&	'	(	)	*	+	,	-	.	/
30	31	32	33	34	35	36	37	38	39	3A	3B	3C	3D	3E	3F
	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>
40	41	42	43	44	45	46	47	48	49	4A	4B	4C	4D	4E	4F
	Ю	а	б	ц	д	е	ф	г	х	и	й	к	л	м	н
50	51	52	53	54	55	56	57	58	59	5A	5B	5C	5D	5E	5F
	п	я	р	с	т	у	ж	в	ь	ы	з	ш	э	щ	ч
60	61	62	63	64	65	66	67	68	69	6A	6B	6C	6D	6E	6F
	Ю	А	Б	Ц	Д	Е	Ф	Г	Х	И	Й	К	Л	М	Н
70	71	72	73	74	75	76	77	78	79	7A	7B	7C	7D	7E	7F
	П	Я	Р	С	Т	У	Ж	В	Ь	Ы	З	Ш	Э	Щ	Ч

20	21	22	23	24	25	26	27	28	29	2A	2B	2C	2D	2E	2F
	!	"	#	¤	%	&	'	(	)	*	+	,	-	.	/
30	31	32	33	34	35	36	37	38	39	3A	3B	3C	3D	3E	3F
	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>
40	41	42	43	44	45	46	47	48	49	4A	4B	4C	4D	4E	4F
	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N
50	51	52	53	54	55	56	57	58	59	5A	5B	5C	5D	5E	5F
	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^
60	61	62	63	64	65	66	67	68	69	6A	6B	6C	6D	6E	6F
	Ю	А	Б	Ц	Д	Е	Ф	Г	Х	И	Й	К	Л	М	Н
70	71	72	73	74	75	76	77	78	79	7A	7B	7C	7D	7E	7F
	П	Я	Р	С	Т	У	Ж	В	Ь	Ы	З	Ш	Э	Щ	Ч

1-байтовая кодировка

Бит чётности (Parity bit)

Способ контроля целостности данных, широко употребительный в докомпьютерные времена.

В 8-битный байт помещается 7-битный код (напр. ASCII) и добавляется контрольный бит. Его значение (0 или 1) устанавливается так, чтобы общее количество единичных битов в байте было чётным.

Расширение таблицы ASCII

Если не использовать бит четности, таблицу ASCII можно расширить до 256 символов, добавив еще 128 символов с кодами $80_{16} \dots FF_{16}$ (128 ... 255).

1-байтовая кодировка

Расширение ASCII CP437 (Codepage 437)

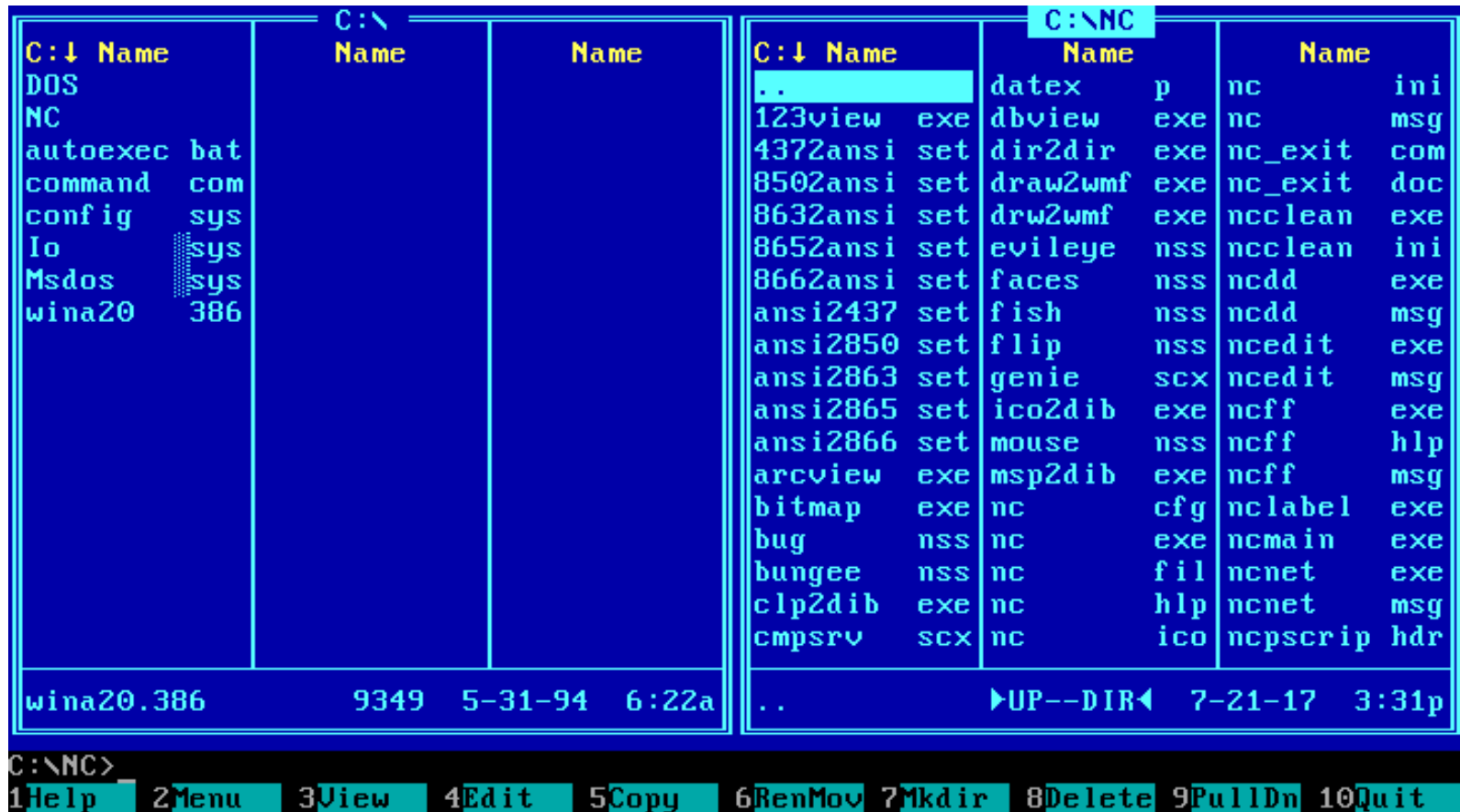
Кодовая страница,
использовавшаяся
в первых IBM PC (1981 г.)

В ПЗУ видеоадаптера был
зашит шрифт с набором
символов для поддержки
некоторых национальных
алфавитов (Ç ü é à Ñ...) и
псевдографики

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0		␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣
1	▶	◀	†	‡	§	¶	⌘	⌚	↑	↓	→	←	⌞	⌟	⌠	⌡
2		!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
3	␣	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	␣	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	␣	P	Q	R	S	T	U	V	W	X	Y	Z	[	]	^	_
6	␣	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	␣	p	q	r	s	t	u	v	w	x	y	z	{	}	~	␣
8	␣	ç	ü	é	â	ä	å	ç	ê	ë	ö	ï	í	î	ï	ä
9	␣	É	á	í	ó	ñ	ñ	ú	û	ü	ö	ï	ç	½	¾	À
A	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣
B	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣
C	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣
D	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣
E	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣
F	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣

1-байтовая кодировка

Псевдографика в текстовом режиме DOS (Norton Commander)



8-битная кодировка кириллицы

Кириллизация расширенного ASCII: KOI-8 (UNIX)

Распространенные версии: **KOI-8R** (только русский);
KOI-8U (с поддержкой букв украинского алфавита – на рисунке);

80	81	82	83	84	85	86	87	88	89	8A	8B	8C	8D	8E	8F
—		Г	г	Л	л	Т	т	Т	т	+	■	■	■	■	■
90	91	92	93	94	95	96	97	98	99	9A	9B	9C	9D	9E	9F
⋯	⋯	⋯	Г	■	●	√	≈	≤	≥		↓	°	2	.	÷
A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF
=		Г	ё	е	Г	і	ї	Г	Г	Г	Г	Г	Г	Г	Г
B0	B1	B2	B3	B4	B5	B6	B7	B8	B9	BA	BB	BC	BD	BE	BF
		Г	Ё	Е	Г	І	Ї	Г	Г	Г	Г	Г	Г	Г	©
C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	CA	CB	CC	CD	CE	CF
Ю	А	Б	Ц	Д	Е	Ф	Г	Х	И	Й	К	Л	М	Н	О
D0	D1	D2	D3	D4	D5	D6	D7	D8	D9	DA	DB	DC	DD	DE	DF
П	Я	Р	С	Т	У	Ж	В	Ь	Ы	З	Ш	Э	Щ	Ч	Ъ
E0	E1	E2	E3	E4	E5	E6	E7	E8	E9	EA	EB	EC	ED	EE	EF
Ю	А	Б	Ц	Д	Е	Ф	Г	Х	И	Й	К	Л	М	Н	О
F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	FA	FB	FC	FD	FE	FF
П	Я	Р	С	Т	У	Ж	В	Ь	Ы	З	Ш	Э	Щ	Ч	Ъ

KOI-RU (рус., укр. и бел.: буквы Ё, Ё на позициях AE, BE)

Есть версии для сербского, чувашского, таджикского и др. языков

8-битная кодировка кириллицы

КОИ-8 применяется в ОС UNIX и ее клонах.
Эту кодировку используют многие веб-сайты.

Символы кириллицы помещены во второй половине кодовой таблицы таким образом, что их позиции соответствуют фонетическим аналогам латиницы из первой части таблицы.

Если в тексте, написанном в КОИ-8, для каждого символа убрать старший бит, то получится относительно читаемый текст. Напр., «мама мыла раму» превратится в «MAMA MYLA RAMU».

Из-за этого в КОИ-8 символы кириллицы расположены не в алфавитном порядке.

8-битная кодировка кириллицы

«Основная» кодовая таблица

Применялась в ЕС-1841 (Минск), Искра-1030 (Смоленск, Курск), Поиск-2 (Киев) и др. (1986-1994)

Не прижилась из-за несовместимости с псевдографикой CP437

±	Ш	ґ	ҕ		п	ғ	π	ц	Е	Р	ш	Ј	ґ		п
Г	г	Ј	Љ	=		ґ		±			▨	▩	▪		±
г	г	Ј	Љ	—		т	†	⊥	†	†	■	■	■	■	■
А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
Ё	ё	‘	’	‘	’	>	<	↑	↓	÷	±	№	¤	■	nbsp

8-битная кодировка кириллицы

Вид программы Norton Commander в «основной» кодировке

The screenshot shows the Norton Commander interface with two panels. The left panel displays the directory structure of drive A:\, and the right panel displays the contents of drive B:\. The interface is in Russian, with Cyrillic text for file names and extensions. The bottom of the screen shows a menu bar with options: 1 Help, 2 Menu, 3 View, 4 Edit, 5 Copy, 6 RenMov, 7 Mkdir, 8 Delete, 9 PullDn, 10 Quit.

Left	Files	Commands	Options	Right
A:\ 36333333336333 B:\ 3633333333633333				
K	Name	Size	Date	Time
KDOS	Γ	Γ	Γ	Γ
Kio	AsysΓ	48774Γ	8-19-94Γ	1:00pKKr8010m
Knsdos	AsysΓ	38138Γ	8-19-94Γ	1:00pKKrus
Kautoexec	batΓ	255Γ	5-26-11Γ	5:07pKKsystem
Kcommand	conΓ	54869Γ	8-19-94Γ	1:00pKKtbl
Kconfig	sysΓ	150Γ	5-26-11Γ	4:54pKKvdisk
Krus	exeΓ	12084Γ	11-01-89Γ	12:00aKKATPMB
Kuimage	exeΓ	8494Γ	10-05-84Γ	10:25aKKΓPAΦHKA
K	Γ	Γ	Γ	KKΓPAΦTABI
K	Γ	Γ	Γ	KKДИСК
K	Γ	Γ	Γ	KKДИСКОП
K	Γ	Γ	Γ	KKДОС
K	Γ	Γ	Γ	KKДОС
K	Γ	Γ	Γ	KKДСРАВН
K	Γ	Γ	Γ	KKЗАП
K	Γ	Γ	Γ	KKКОЛОБОК
K	Γ	Γ	Γ	KKMB6
B:\ 3633333333633333				
Kuimage.exe	8494	10-05-84	10:25a	KKКОЛОБОК.KHД
B:\>				
1Help 2Menu 3View 4Edit 5Copy 6RenMov 7Mkdir 8Delete 9PullDn 10Quit				

8-битная кодировка кириллицы

Таблица по стандарту ISO 8859-5

Разработана на базе «основной» таблицы. Содержит символы алфавитов: рус., укр., бел. сербск., македонск., болг.

Популярна в Сербии и Македонии.
У нас и в Болгарии не прижилась.

A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF
	Ё	Ђ	Ѓ	Є	Ѕ	І	Ї	Ј	Љ	Њ	Ћ	Ќ	-	Ў	Џ
B0	B1	B2	B3	B4	B5	B6	B7	B8	B9	BA	BB	BC	BD	BE	BF
	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О
C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	CA	CB	CC	CD	CE	CF
	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю
D0	D1	D2	D3	D4	D5	D6	D7	D8	D9	DA	DB	DC	DD	DE	DF
	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о
E0	E1	E2	E3	E4	E5	E6	E7	E8	E9	EA	EB	EC	ED	EE	EF
	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю
F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	FA	FB	FC	FD	FE	FF
	№	ё	ђ	ѓ	є	ѕ	і	ї	ј	љ	њ	ћ	ќ	ѕ	џ

8-битная кодировка кириллицы

Болгарская кодировка МИК («Interprog») (М??? Извънстандартна Кодировка)

Применялась на болгарских компьютерах (напр., «Правец») и в болгарской локализации ОС UNIX.

Приводим ради полноты картины.

80	А	81	Б	82	В	83	Г	84	Д	85	Е	86	Ж	87	З	88	И	89	Й	8A	К	8B	Л	8C	М	8D	Н	8E	О	8F	П
90	Р	91	С	92	Т	93	У	94	Ф	95	Х	96	Ц	97	Ч	98	Ш	99	Щ	9A	Ъ	9B	Ы	9C	Ь	9D	Э	9E	Ю	9F	Я
A0	а	A1	б	A2	в	A3	г	A4	д	A5	е	A6	ж	A7	з	A8	и	A9	й	AA	к	AB	л	AC	м	AD	н	AE	о	AF	п
B0	р	B1	с	B2	т	B3	у	B4	ф	B5	х	B6	ц	B7	ч	B8	ш	B9	щ	BA	ъ	BB	ы	BC	ь	BD	э	BE	ю	BF	я
C0	Љ	C1	Њ	C2	Ћ	C3	Ќ	C4	—	C5	†	C6	‡	C7	‖	C8	ℓ	C9	ѓ	CA	џ	CB	т	CC	ђ	CD	=	CE	†	CF	Џ
D0	Џ	D1	Џ	D2	Џ	D3		D4	†	D5	№	D6	§	D7	¶	D8	Ј	D9	Ј	DA	г	DB	■	DC	■	DD	■	DE	■	DF	■
E0	α	E1	β	E2	Г	E3	π	E4	Σ	E5	σ	E6	μ	E7	τ	E8	Φ	E9	θ	EA	Ω	EB	δ	EC	∞	ED	∅	EE	ε	EF	п
F0	≡	F1	±	F2	≥	F3	≤	F4	↑	F5	↓	F6	÷	F7	∞	F8	°	F9	.	FA	.	FB	√	FC	n	FD	2	FE	■	FF	

8-битная кодировка кириллицы

«Альтернативная» кодовая таблица, CP866

Введена в практику Microsoft при локализации MS DOS.
Осн. преимущество: совместимость с псевдографикой CP437.

CP 866ukr – нестандартная модификация
с «І», «і» вместо «У», «у» («Г», «г» - отсутствуют.)

80	А	81	Б	82	В	83	Г	84	Д	85	Е	86	Ж	87	З	88	И	89	Й	8A	К	8B	Л	8C	М	8D	Н	8E	О	8F	П
90	Р	91	С	92	Т	93	У	94	Ф	95	Х	96	Ц	97	Ч	98	Ш	99	Щ	9A	Ъ	9B	Ы	9C	Ь	9D	Э	9E	Ю	9F	Я
A0	а	A1	б	A2	в	A3	г	A4	д	A5	е	A6	ж	A7	з	A8	и	A9	й	AA	к	AB	л	AC	м	AD	н	AE	о	AF	п
B0	▒	B1	▒	B2	▒	B3		B4		B5		B6		B7		B8		B9		BA		BB		BC		BD		BE		BF	
C0	┐	C1	┐	C2	┐	C3	┐	C4	┐	C5	┐	C6	┐	C7	┐	C8	┐	C9	┐	CA	┐	CB	┐	CC	┐	CD	┐	CE	┐	CF	┐
D0	└	D1	└	D2	└	D3	└	D4	└	D5	└	D6	└	D7	└	D8	└	D9	└	DA	└	DB	└	DC	└	DD	└	DE	└	DF	└
E0	р	E1	с	E2	т	E3	у	E4	ф	E5	х	E6	ц	E7	ч	E8	ш	E9	щ	EA	ъ	EB	ы	EC	ь	ED	э	EE	ю	EF	я
F0	Ё	F1	ё	F2	Є	F3	е	F4	İ	F5	ï	F6	Û	F7	ұ	F8	°	F9	•	FA	·	FB	√	FC	№	FD	¤	FE	■	FF	

8-битная кодировка кириллицы

MacUkrainian

(= MacCyrillic + «Ґ», «ґ» на месте «ѣ», «Ѧ»)

Применяется только на компьютерах Apple Macintosh

80	А	81	Б	82	В	83	Г	84	Д	85	Е	86	Ж	87	З	88	И	89	Й	8A	К	8B	Л	8C	М	8D	Н	8E	О	8F	П
90	Р	91	С	92	Т	93	У	94	Ф	95	Х	96	Ц	97	Ч	98	Ш	99	Щ	9A	Ъ	9B	Ы	9C	Ь	9D	Э	9E	Ю	9F	Я
A0	†	A1	°	A2	Ґ	A3	£	A4	§	A5	•	A6	¶	A7	І	A8	®	A9	©	AA	™	AB	Ђ	AC	Ѓ	AD	Ѕ	AE	Ї	AF	Ї
B0	∞	B1	±	B2	≤	B3	≥	B4	і	B5	μ	B6	ґ	B7	Ј	B8	Є	B9	е	BA	Ї	BB	ї	BC	Љ	BD	љ	BE	Њ	BF	њ
C0	j	C1	S	C2	¬	C3	√	C4	f	C5	≈	C6	Δ	C7	«	C8	»	C9	...	CA		CB	Ђ	CC	Ѓ	CD	Ѕ	CE	Ї	CF	S
D0	-	D1	-	D2	“	D3	”	D4	‘	D5	,	D6	÷	D7	”	D8	Ї	D9	Ї	DA	Љ	DB	љ	DC	Њ	DD	Ѕ	DE	Ї	DF	я
E0	а	E1	б	E2	в	E3	г	E4	д	E5	е	E6	ж	E7	з	E8	и	E9	й	EA	к	EB	л	EC	м	ED	н	EE	о	EF	п
F0	р	F1	с	F2	т	F3	у	F4	ф	F5	х	F6	ц	F7	ч	F8	ш	F9	щ	FA	ъ	FB	ы	FC	ь	FD	э	FE	ю	FF	ѧ

8-битная кодировка кириллицы

Windows-1251 (CP1251)

Присутствуют все обычные типографские символы (кроме значка удара); содержит символы алфавитов: рус., укр., бел. сербск., македонск., болг.

Отсутствуют символы псевдографики, не нужные для Windows.

80	Ѓ	Ѕ	,	Ѕ	...	†	‡	€	%	Љ	<	Њ	Ќ	Ѕ	Љ
90	ђ	ѐ	ѐ	ѐ	•	-	-		™	љ	>	њ	ќ	ћ	џ
A0	Ў	Ў	Ј	Њ	Г	І	Ѕ	Ѓ	©	Є	«	¬	-	®	Ї
B0	°	±	І	і	Г	μ	¶	•	ё	№	е	»	ј	ѕ	ї
C0	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О
D0	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю
E0	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о
F0	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю

8-битная кодировка кириллицы

EBCDIC (Extended Binary Coded Decimal Information Code)

Несовместимая с ASCII латинская кодировка. Активно использовалась до 1990-х годов для мейнфреймов IBM и их советских клонов ЕС ЭВМ. Остаётся основной в z/OS для современных мейнфреймов IBM Z.

Кириллическое расширение – **ДКОИ-8**
(места размещения русских букв показаны желтым фоном).

hex	x0	x1	x2	x3	x4	x5	x6	x7	x8	x9	xA	xB	xC	xD	xE	xF
0x	NUL	SOH	STX	ETX	PF	HT	LC	DEL			SMM	VT	FF	CR	SO	SI
1x	DLE	DC1	DC2	TM	RES	NL	BS	IL	CAN	EM	CC	CU1	IFS	IGS	IRS	IUS
2x	DS	SOS	FS		BYP	LF	ETB	ESC			SM	CU2		ENQ	ACK	BEL
3x			SYN		PN	RS	UC	EOT				CU3	DC4	NAK		SUB
4x	SP										d	.	<	(	+	
5x	&										!	\$	*	)	;	^
6x	-	/										!	%	'	>	?
7x											:	#	@	^	=	"
8x		a	b	c	d	e	f	g	h	i						
9x		j	k	l	m	n	o	p	q	r						
Ax		~	s	t	u	v	w	x	y	z						
Bx																
Cx	{	A	B	C	D	E	F	G	H	I						
Dx	}	J	K	L	M	N	O	P	Q	R						
Ex	\		S	T	U	V	W	X	Y	Z						
Fx	0	1	2	3	4	5	6	7	8	9						

8-битная кодировка кириллицы

Пример неправильно выбранной кодировки

Windows-1251 (CP 1251)

Рученьки терпнуть, злипаються віченьки...
Боже, чи довго тягти?
З раннього ранку до пізньої ніченьки
Голкою денно верти.

KOI-8R

псвЕМЗЙХ РЕПОМСРЭ, ГКХОЮЧРЭЯЬ БЁВЕМЗЙХ...
аНФЕ, ВХ ДНБЦН РЬЦРХ?
г ГЮММЭНЦН ГЮМЙС ДН ОЁГМЭНЮ МЁВЕМЗЙХ
цНКЙНЧ ДЕММН БЕПРХ.

«Альтернативная» (CP866)

└єЎхэЊш ЄхЁязєЄЊ, чышяр■ЄЊё т | ўхэЊш...
└юцх, ўш фютую Є уЄш?
└ ЁрээЊую Ёрээє фю я | чэЊю┐ э | ўхэЊш
└юью■ фхээю тхЁєш.

MacCyrillic

-ученьки терпнуть, злипаються в≥ченьки...
Soже, чи довго тѧгти?
« раннього ранку до п≥зньої н≥ченьки
√олкою денно верти.

8-битная кодировка кириллицы

Основные кириллические кодировки:

- КОИ-8
- ISO 8859-5 (и ее вариант «Основная»)
- Болгарская МИК
- CP866 («Альтернативная»)
- Mac
- CP1251 (Windows-1251)
- ДКОИ-8

С китайскими иероглифами ситуация так же запутана?

8-битная кодировка кириллицы

- <https://ru.wikipedia.org/wiki/ASCII>
- <http://book.itep.ru/10/ascii.htm>
- https://ru.wikipedia.org/wiki/Бит_чётности
- <https://ru.wikipedia.org/wiki/CP437>
- <http://www.inp.nsk.su/~baldin/Cyrillic-HOWTO-russian/ch03.html>

Unicode

Проблемы из-за множественности кодировок:

- **неправильная раскодировка («кракозябры»).**
Решение: либо указание кодировки, либо внедрение общей для всех языков «широкой» кодировки.
- **Ограниченность набора символов.**
Решение: либо переключение шрифтов внутри документа, либо внедрение «широкой» кодировки.
- **Преобразование одной кодировки в другую.**
Решение: либо таблицы перекод. для каждой пары кодировок, либо промежуточное преобразования в «широкую» кодировку.
- **Дублирование шрифтов** (для каждой кодировки – свой шрифт, даже если часть символов совпадает).
Решение: введение единого реестра символов и создание «больших» шрифтов, из которых впоследствии выбирать нужные для данной кодировки символы.

Unicode



Unicode – стандарт кодирования символов для **всех** письменных языков (в т.ч. мертвых).

Универсальный набор символов (Universal character set, UCS) задает однозначное соответствие символов и кодов.

Первые 128 символов соответствуют ASCII.

Стандартная таблица UCS постоянно расширяется. По состоянию на март 2020 она кодирует **154 алфавита** и содержит **143 859 символов**. Имеются также математические, технические, музыкальные символы, пиктограммы...

Форматы представления юникода (Unicode Transformation Format, UTF):

- UTF-8,
- UTF-16,
- UTF-32.

Unicode



UTF-16

Первая версия Юникода (1991) ограничивалась $2^{16} = 65\,536$ (64k) символами $0000_{16} \dots FFFF_{16}$ и использовала 16-битную кодировку.

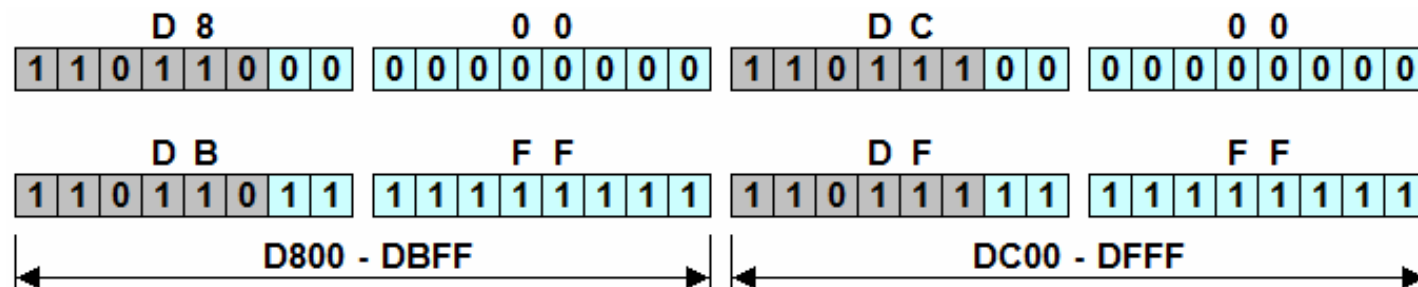
Во второй версии Юникода (1996) кодовая область расширена. Для совместимости с предыдущей версией символы в диапазонах $0000_{16} \dots D7FF_{16}$ (первые 54k) и $E000_{16} \dots FFFF_{16}$ (последние 8k) кодируются как и ранее, 16-битным словом.

Исключенный диапазон $D800_{16} \dots DFFF_{16}$ длиной $2k = 2^{11} = 2048$ кодов, ранее принадлежавший к области «символов для частного использования», используется для т. наз. **«суррогатных пар»** – символов, которые кодируются **двумя** 16-битными словами.

Unicode

UTF-16: «Суррогатная пара»

1-е 16-битное слово начинается с 110110_2 , а 2-е – с 110111_2 .
Остальные 20 бит «суррогатной пары» могут использоваться для кодирования еще 2^{20} символов.



Т. обр, общее число символов, доступных для кодирования в UTF-16, составляет $2^{16} - 2^{11} + 2^{20} = 17 \times 64k - 2k = 1\,112\,064$ символа.

Unicode



Кодовое пространство Unicode

разбито на **17 плоскостей (planes)** по $2^{16} = 65\,536$ (64k) символов.

Нулевая плоскость наз. **базовой** и содержит символы всех употребительных письменностей. Её символы (кроме 2k кодов, которые не используются) кодируются одним 16-битным словом.

Остальные плоскости (от 1 до 16) наз. **дополнительными**, их символы в диапазоне $10000_{16} \dots 10FFFF_{16}$ (больше 16 бит) кодируются по следующей схеме:

Из кода символа вычитается 10000_{16} . В результате получится значение от 00000_{16} до $FFFFF_{16}$, которое помещается в разрядную сетку 20 бит.

Дополнительные плоскости используются для исторических письменностей, редких и архаичных китайских иероглифов...

Плоскости 15 и 16 предназначены для частного употребления.

Unicode

Фрагмент таблицы
символов Unicode
с кодами **0400 ... 04FF**,
содержащий символы
кириллицы

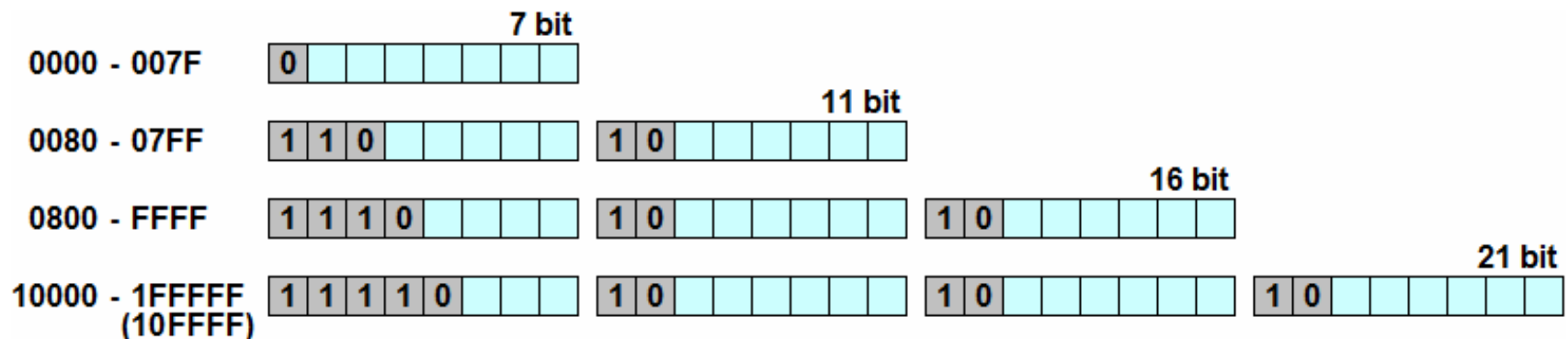
	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
	Ё	Ђ	Ѓ	Є	Ѕ	І	Ї	Ј	Љ	Њ	Ћ	Ќ		Ў	Ў
10	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О
20	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю
30	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о
40	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю
	ё	ђ	ѓ	є	ѕ	і	ї	ј	љ	њ	ћ	ќ		ў	џ
60	Ѡ	ѡ	Ѣ	ѣ	Ѥ	ѥ	Ѧ	ѧ	Ѩ	ѩ	Ѫ	ѫ	Ѭ	ѭ	Ѯ
70	Ѱ	ѱ	Ѳ	ѳ	Ѵ	ѵ	Ѷ	ѷ	Ѹ	ѹ	Ѻ	ѻ	Ѽ	ѽ	Ѿ
80	Ѽ	ѽ	Ѿ												
90	Г	Г	г	г	Б	Б	Ж	Ж	З	З	К	К	К	К	К
A0	К	к	Н	н	Н	н	Љ	Љ	Ќ	Ќ	Ћ	Ћ	Ќ	Ќ	Ќ
B0	Ѵ	ѵ	Х	х	Ц	ц	Ч	ч	Ч	ч	Ѡ	ѡ	Ѣ	ѣ	Ѥ
C0	І	Ж	Ж	К	К			Н	Н			Ч	Ч		
D0	Ѐ	Ё	Ђ	Ѓ	Є	Ѕ	І	Ї	Ї	Љ	Њ	Ћ	Ќ	Ў	Ў
E0	З	З	Й	Й	Й	Й	Ќ	Ќ	Ќ	Ќ	Ќ	Ќ	Ќ	Ќ	Ќ
F0	Ў	Ў	Ў	Ў	Ў	Ў			Ы	Ы					

Unicode

UTF-8: компактное кодирование латиницы

Если текст состоит только из символов с номерами меньше 128, то при записи в UTF-8 он превращается в обычный текст ASCII.

Остальные символы Unicode изображаются кодами от 2 до 4 байт, где маска первого байта – 11xxxxxx, а остальных – 10xxxxxx.



Хотя один и тот же символ можно указать несколькими способами, только самый короткий из них правильный. Для совместимости с UTF-16 диапазон символов ограничен кодом $10FFFF_{16}$.

Unicode



UTF-32

Использует для кодирования любого символа ровно 32 бита.

Преимущество перед кодировками переменной длины – лёгкость получения n -ой кодовой позиции (напр., 1000-й символ занимает 4 байта 3997...4000). Напротив, коды с переменной длиной требуют последовательного доступа к n -ой кодовой позиции.

Недостаток – неэффективное использование памяти. Символы, лежащие за пределами нулевой (базовой) плоскости, редко используются в большинстве текстов. Поэтому удвоение, по сравнению с UTF-16, пространства памяти обычно не оправдано.

Не рекомендуется для применения в HTML.

Unicode

- <https://ru.wikipedia.org/wiki/Юникод>
- [https://ru.wikipedia.org/wiki/Плоскость_\(Юникод\)](https://ru.wikipedia.org/wiki/Плоскость_(Юникод))
- <http://foxtools.ru/Unicode>
- <https://unicode.org/charts/>
- <https://gtavrl.ru/skolko-simvolov-v-tablice-unicode-ispolzovanie-yunikoda-s-font-face-ikonkami/>

RTF – Rich Text Format

Rich Text Format, RTF («формат обогащённого текста») – **межплатформенный** формат хранения текстовых документов с форматированием. Предложен в 1982 как формат с управляющими метками («тэгами») для редактора Word.

```
{\rtf1\ansi\deff0{\fonttbl{\f0\fnil\charset0 Courier New;}}
{\*generator Msftedit 5.41.21.2510;}\viewkind4\uc1\pard\lang1033\f0\fs22 Jingle
bells, jingle bells\par
Jingle all the way\par
Oh, what fun it is to ride\par
In a one horse open sleigh\par
}
```

Текст в формате RTF обычно кодируется 7-битными символами. Остальные символы могут кодироваться двумя способами:

- либо кодами в кодировке символов указанного языка,
- либо кодами в Юникоде.

RTF – Rich Text Format

Напр., если задан язык **lang1058** (укр.), то ему соответствует кодировка Windows-1251, а коду **\'d0** соответствует буква **Р**.

```
{\rtf1\ansi\deff0{\fonttbl{\f0\fnil\fcharset204{*\fname Courier New;}Courier New
CYR;}}{\f1\fnil\fcharset0 Courier New;}}
{\*\generator Msftedit 5.41.21.2510;}\viewkind4\uc1\pard\lang1058\f0\fs22\'d0\'f3
\'f7\'e5\'ed\'fc\'ea\'e8 \'f2\'e5\'f0\'ef\'ed\'f3\'f2\'fc,
\'e7\'eb\'e8\'ef\'e0\'fe\'f2\'fc\'f1\'ff \'e2\'b3\'f7\'e5\'ed\'fc\'ea\'e8...\par
\'c1\'ee\'e6\'e5, \'f7\'e8 \'e4\'ee\'e2\'e3\'ee \'f2\'ff\'e3\'f2\'e8?\par
\'c7 \'f0\'e0\'ed\'ed\'fc\'ee\'e3\'ee \'f0\'e0\'ed\'ea\'f3 \'e4\'ee
\'ef\'b3\'e7\'ed\'fc\'ee\'bf \'ed\'b3\'f7\'e5\'ed\'fc\'ea\'e8\par
\'c3\'ee\'eb\'ea\'ee\'fe \'e4\'e5\'ed\'ed\'ee \'e2\'e5\'f0\'f2\'e8.\par
\par
\lang1033\f1\par
}
```

Если требуется символ в Юникоде, то используется префикс **\u**, и сразу после него – 16-битное число в 10-ной системе.

RTF – Rich Text Format

- https://ru.wikipedia.org/wiki/Rich_Text_Format
- <https://filesreview.com/ru/info/rtf>
- https://windows-school.ru/blog/chto_predstavljajut_soboj_rtf/2018-08-06-156
- <https://www.science.co.il/language/Locale-codes.php>

Компьютерные шрифты

Шрифт – это набор визуально различных, но единообразно спроектированных знаков всего алфавита, включая буквенные, цифровые и прочие символы.

Компьютерный шрифт – это файл, содержащий в себе описание данного набора символов, используемый для их отображения программой или операционной системой.

По принципу начертания букв компьютерные шрифты делятся на два типа:

- **растровые**, где каждый символ описан в виде набора пикселей, расположенных в узлах сетки растра (т.е. как точечный рисунок);
- **векторные**, где каждый символ имеет математическое описание и представляется в виде набора точек (векторов) и соединяющих их линий.

Компьютерные шрифты

Растровые шрифты немасштабируемы. Они непригодны для высококачественной печати и используются в основном в консоли и в программах с текстовым интерфейсом.

Широко использовались в эпоху матричных принтеров и символьных мониторов.

Файлы Windows, содержащие растровые шрифты, имеют расширение **.FON** (Font).

Распpовыѝ шpифт

Компьютерные шрифты

В **векторных шрифтах** каждый знак описан с помощью координат опорных точек, которые соединены прямыми или кривыми и образуют контур знака без привязки к абсолютному размеру или разрешению.

Векторные шрифты одинаково выглядят как на экране, так и на бумаге, поскольку масштабирование символов производится простым умножением имеющихся координат на необходимый множитель.

В Windows применяются шрифты TrueType, где опорные точки соединяются кривыми 2-го порядка. Файлы, их содержащие, имеют расширение **.TTF** (TrueType Font).

Векторный шрифт

Характеристики шрифтов

Гарнитура – набор шрифтов с общими стилевыми признаками (дизайном), имеющих различные начертания и размеры (кегли).

Кегль – размер шрифта, включая нижние и верхние выносные элементы букв или знаков. Измеряется в типографских **пунктах** (points). 1 пункт = 1/72 дюйма = 0,353 мм.



Начертание — набор шрифтов всех размеров определенной насыщенности и стиля. Как правило, шрифт имеет минимум четыре начертания: **нормальный**, **полужирный (bold)**, *курсив (italic)* и **полужирный курсив (bold italic)**.

Шрифт — набор символов определенного размера и начертания (напр., полужирный шрифт Arial размером 12 пунктов).

Характеристики шрифтов

Объём Луны — $2,196 \times 10^{10} \text{ км}^3$ | 36 pt

Одинаковый кегль в
различных шрифтах

Гарнитура Futura PT
в разных начертаниях

Futura PT

ШРИФТОВАЯ СИСТЕМА

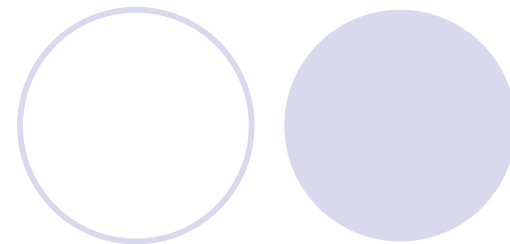
в 22 начертаниях

включающая четырнадцать нормальных

И ВОСЕМЬ УЗКИХ

начертаний

Пропорциональные и моноширинные шрифты



В **пропорциональных** шрифтах ширины знаков варьируются в зависимости от конструкции букв (в отличие от моноширинного), поэтому, например, литер «і» в одной строке поместится больше, чем букв «щ» того же кегля.

Моноширинные шрифты имитируют литеры пишущих машинок, матричных принтеров и символьных терминалов.

В этих шрифтах ширина всех символов одинакова.

Применяются для набора таблиц, листинга программ и в других случаях, когда требуется полное совпадение колонок и знаков по вертикали набора.

Текст, записанный моноширинным шрифтом, выглядит и читается хуже.

Пропорциональные и моноширинные шрифты

Шрифты **Times** (засечный) и **Arial** (рубленый) – пропорциональные.

Шрифты **Courier** (засечный), **Consolas** и **Fixedsys** (рубленые) – моноширинные.



Пример векторных
шрифтов обоих типов

ABCDEFGHIJKLMN
NOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
nopqrstuvwxyz
0123456789!/?#
%&* ((/ \) }

Пример моноширинного
растрового 5×9 шрифта

Стилистические группы шрифтов

Основные стилистические группы шрифтов:

- засечные (серифные, serif);
- рубленые или гротески (без засечек, sans-serif);
- акцидентные (декоративные), в т. ч. рукописные;
- символные и орнаментальные.

Стилистические группы шрифтов

Засечные шрифты имеют **засечки** (серифы) – поперечные элементы на концах штрихов буквы. Они делают шрифт более четким, зрительно «связывая» слово в одно целое, и тем самым помогают читать текст.

Засечные шрифты единственно приемлемые при печати больших объемов текста в книгах и журналах.

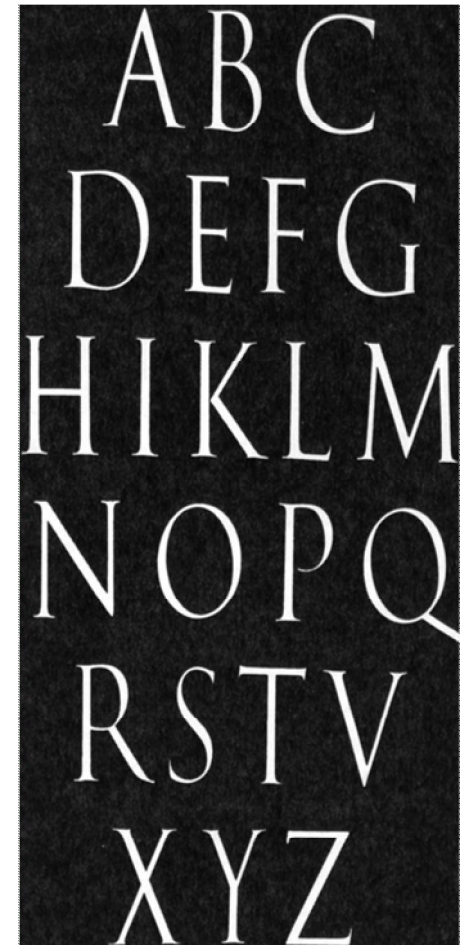
Наиболее популярный засечный шрифт в Windows – **Times New Roman**



Стилистические группы шрифтов

Впервые засечки на буквах применили древние римляне, а первый подобный типографский шрифт создали итальянские первопечатники в 1460-х годах.

OldStyle
Courier New
EDITION
Century
Cairo



Стилистические группы шрифтов

Рубленые шрифты или гротески не имеют засечек и единообразны по толщине. Применяются для малых объемов текста – заголовков, рекламных вывесок, дорожных указателей.

Наиболее популярный рубленый шрифт в Windows – **Arial**.

Первые типографские шрифты без засечек появились в 1810-х годах в рекламных изданиях в Англии.

Arial

BerlinEmail

GeosansLight

Duepuntozero

headthinker

Стилистические группы шрифтов

Акциденция (от лат. accidentia – случай, случайность) – это малообъемная печатная продукция (объявления, афиши, грамоты, календари и др.), выполняемая посредством акцидентного, т.е. особо сложного набора, для которого применяются специальные акцидентные и декоративные шрифты, орнаменты, линейки, особые приемы верстки и т.п.

Акцидентные (декоративные) шрифты

– самая большая по количеству шрифтов категория. Они предназначены для заголовков и других небольших отрывков текста для привлечения и акцентирования внимания.



Стилистические группы шрифтов

Примеры акцидентных шрифтов



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

ТРИПІАЛЯ

A B C D E F	А Б В Г Г Д Е
G H I J K L	Є Ж З Й І К
M N O P Q	Л М Н О П Р
R S T U V W	С Т У Ф Х Ц
X Y Z . ! ? ; -	Ч Ш Щ Ъ Ы Ь
0 1 2 3 4 5 6 7 8 9	Э Ю Я . ! ? ; :

Стилистические группы шрифтов

Часто в акцидентных шрифтах выделяют отдельную подгруппу – **рукописные шрифты**, которые имитируют текст, написанный вручную ручкой, пером, кисточкой и т.п.

Pushkin

Vivaldi

Pristina

Lucida Handwriting

Henry Morgan Hand

Набор кириллических шрифтов

Набор кириллических шрифтов

Набор кириллических шрифтов

Набор кириллических шрифтов

Набор кириллических шрифтов

Стилистические группы шрифтов

Символьные и орнаментальные шрифты.

К этой группе относятся комплекты специальных знаков и наборных украшений.

Наиболее популярный символьный шрифт в Windows – **Wingdings**.



Стилистические группы шрифтов

- <http://www.quickandlazy.appspot.com/>
- [http://rus.paratype.ru/catalog-chapter/a?_=_](http://rus.paratype.ru/catalog-chapter/a?_=)