

# Локализация и интернационализация программного обеспечения

## Инструментарий Современного Программиста

Иван Трепаков

NSU Sys.Pro

# ASCII

## American Standard Code for Information Interchange

```
$ ascii -d
```

0 NUL	16 DLE	32	48 0	64 @	80 P	96 `	112 p
1 SOH	17 DC1	33 !	49 1	65 A	81 Q	97 a	113 q
2 STX	18 DC2	34 "	50 2	66 B	82 R	98 b	114 r
3 ETX	19 DC3	35 #	51 3	67 C	83 S	99 c	115 s
4 EOT	20 DC4	36 \$	52 4	68 D	84 T	100 d	116 t
5 ENQ	21 NAK	37 %	53 5	69 E	85 U	101 e	117 u
6 ACK	22 SYN	38 &	54 6	70 F	86 V	102 f	118 v
7 BEL	23 ETB	39 '	55 7	71 G	87 W	103 g	119 w
8 BS	24 CAN	40 (	56 8	72 H	88 X	104 h	120 x
9 HT	25 EM	41 )	57 9	73 I	89 Y	105 i	121 y
10 LF	26 SUB	42 *	58 :	74 J	90 Z	106 j	122 z
11 VT	27 ESC	43 +	59 ;	75 K	91 [	107 k	123 {
12 FF	28 FS	44 ,	60 <	76 L	92 \	108 l	124
13 CR	29 GS	45 -	61 =	77 M	93 ]	109 m	125 }
14 SO	30 RS	46 .	62 >	78 N	94 ^	110 n	126 ~
15 SI	31 US	47 /	63 ?	79 O	95 _	111 o	127 DEL

# Расширения ASCII

## ASCII

00:  
10:  
20: !"#\$%&'()\*+,-./  
30: 0123456789:;<=>?  
40: @ABCDEFGHIJKLMNO  
50: PQRSTUVWXYZ[\]^\_  
60: `abcdefghijklmno  
70: pqrstuvwxyz{|}~

# Расширения ASCII




## ASCII

00:	80:
10:	90:
20: <code>!"#\$%&amp;'()*+,-./</code>	a0:
30: <code>0123456789:;&lt;=&gt;?</code>	b0:
40: <code>@ABCDEFGHIJKLMNO</code>	c0:
50: <code>PQRSTUVWXYZ[\]^_</code>	d0:
60: <code>`abcdefghijklmno</code>	e0:
70: <code>pqrstuvwxyz{ }~</code>	f0:

# Расширения ASCII

## ASCII

00: 10: 20: 30: 40: 50: 60: 70:

80: АБВГДЕЖЗИЙКЛМНОП  
90: РСТУФХЦЧШЩЪЫЬЭЮЯ  
a0: абвгдежзийклмноп  
b0:  | | | | | | | | | | | | | | | |  
c0:  | | | | | | | | | | | | | | | |  
d0:  | | | | | | | | | | | | | | | |  
e0: рстуфхцчшщъыьэюя  
f0: ЁёЄєӢӓӮӹ••√№■

## CP866

MS-DOS

## Расширения ASCII

ASCII		CP866		KOI8-R
00:		80:	АБВГДЕЖЗИЙКЛМНОП	-   п Ч Н Т т □ ■ ▩
10:		90:	РСТУФХЦШЩЪЫЬЭЮЯ	▒ ▓ ▒ ■ √ ≈ ≤ ≥ J ° ² ÷
20:	!"#\$%&'()*+,-./	a0:	абвгдежзийклмноп	=    ё ě Ꞥ ꞥ Ꞧ ꞧ Ꞩ ꞩ Ɦ Ɜ Ɡ Ɬ Ɪ ꞯ Ʞ Ʇ Ʝ Ꭓ Ꞵ ꞵ Ꞷ ꞷ Ꞹ ꞹ Ꞻ ꞻ Ꞽ ꞽ Ꞿ ꞿ
30:	0123456789:;<=>?	b0:	▒ ▓ ▒   + = < > ?    Ɦ Ɜ Ɡ Ɬ Ɪ ꞯ Ʞ Ʇ Ʝ Ꭓ Ꞵ ꞵ Ꞷ ꞷ Ꞹ ꞹ Ꞻ ꞻ Ꞽ ꞽ Ꞿ ꞿ	Ё ѐ    Ꞥ ꞥ Ꞧ ꞧ Ꞩ ꞩ Ɦ Ɜ Ɡ Ɬ Ɪ ꞯ Ʞ Ʇ Ʝ Ꭓ Ꞵ ꞵ Ꞷ ꞷ Ꞹ ꞹ Ꞻ ꞻ Ꞽ ꞽ Ꞿ ꞿ ©
40:	@ABCDEFGHIJKLMNO	c0:	Ч Н Т т Ф ф Ц ц Ш ш Щ щ Ъ ъ Ы ы Ь ь Э э Ю ю Я я	ю аб б де ф г х и й к л м н о
50:	PQRSTUVWXYZ[\]^_`	d0:	Ш ш Ч ч Ц ц Ш ш Щ щ Ъ ъ Ы ы Ь ь Э э Ю ю Я я	п я р с т у ж в ы ы з ш э щ ч ь
60:	`abcdefghijklmnopqrstuvwxyz{ }~	e0:	р ст у ф х ц ч ш щ ъ ы ь э ю я	Ю А Б Ц Д Е Ф Г Х И Й К Л М Н О
70:	pqrstuvwxyz{ }~	f0:	Ё ё Є є І і İ i Ÿ ŷ ° • √ № № █	П Я Р С Т У Ж В Ы Ы З Ш Э Щ Ч Ъ

MS-DOS

# Unix

## Расширения ASCII

ASCII	CP866	KOI8-R	CP1251
00:	80: АБВГДЕЖЗИЙКЛМНОП	-   □ ▣ ▤ ▥ ■ █	Ѓ Ѕ , ´ „ „ † € % № < ъ к ђ Ѣ
10:	90: РСТУФХЦЧШЩЪЫЬЭЮЯ	▩ [■ ● √ ≈ ≤ ≥ ] ° ² ÷	ђ ‘ ’ ” • — ™ № > ѝ к ѓ Ѧ
20: _ ! " # \$ % & ' ( ) * + , - . /	a0: абвгдежзийклмноп	=    ф ё г п т ч л ц щ з ж б	Ѹ ѹ Ꙗ Г ¡ § Ğ © « ¬ ~ ® İ
30: 0 1 2 3 4 5 6 7 8 9 ; : = > ?	b0:	©	° ± І і γ μ ¶ · ё № є » ј S s i
40: @ A B C D E F G H I J K L M N O	c0: ¨ ª « ¬ ® ¯ ° ± ¶ · ¸ ¹ º»	ю а б ц д е ф г х и й кл м н о	А Б В Г Д Е Ж З И Й К Л М Н О П
50: P Q R S T U V W X Y Z [\ ] ^ _	d0:                	п я р с т у ж в ы ь з ш э щ ч ь	Р С Т У Ф Х Ц Ч Ш Щ Ъ Ы Ь Э Ю Я
60: ` a b c d e f g h i j k l m n o	e0: р стуфхцчшщъыьэюя	Ю А Б Ц Д Е Ф Г Х И Й К Л М Н О	а б в г д е ж з и й к л м н о п
70: p q r s t u v w x y z {   } ~	f0: ЁёЄєİïıŸ°•●√№▪■	П Я Р С Т У Ж В Ы Ь З Ш Э Щ Ч Ь	р с т у ф х ц ч ш щ ь ь ь э ю я

MS-DOS

# Unix

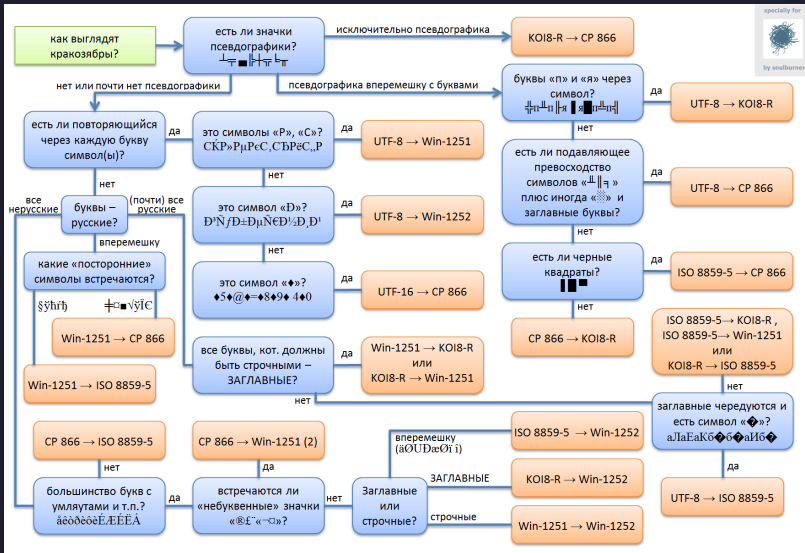
# Windows





оПХБЕР ЛХП!

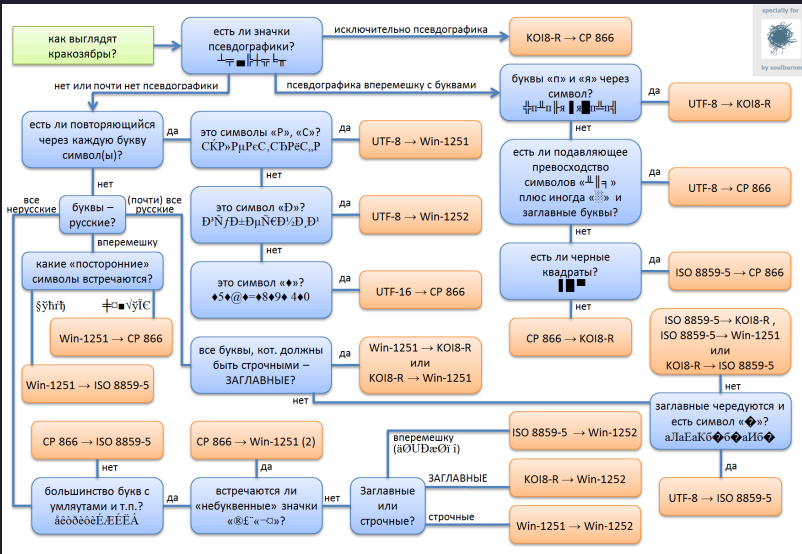
ОПХБЕР ЛХП!



# Крокозябры

ОПХБЕР ЛХП!

```
$ echo "оПХБЕР ЛХП!" \
| iconv -t cp1251 \
| iconv -f koi8-r
Норуаи кую!
```

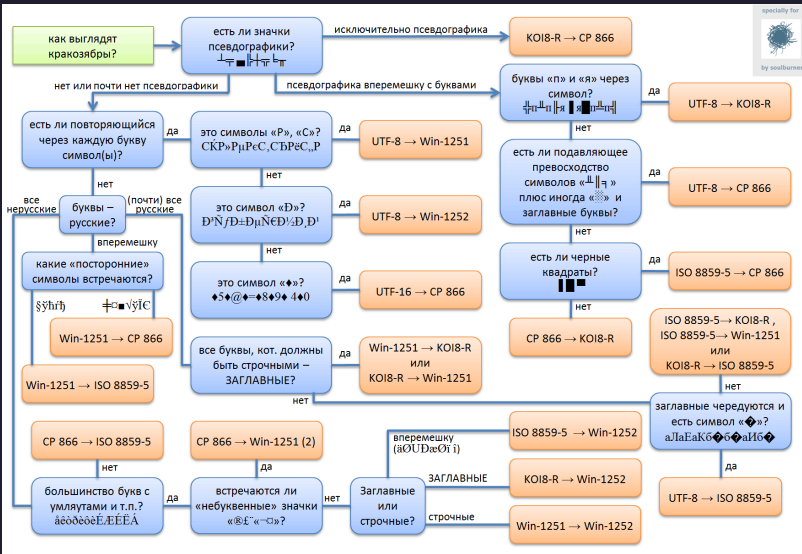


# Крокозябры

ОПХБЕР ЛХП!

```
$ echo "оПХБЕР ЛХП!" \
    | iconv -t cp1251 \
    | iconv -f koi8-r
Норддн куд!
```

```
$ echo "оПХБЕР ЛХП!" \
  | iconv -t koi8-r \
  | iconv -f cp1251
Привет мир!
```



# Крокозябры

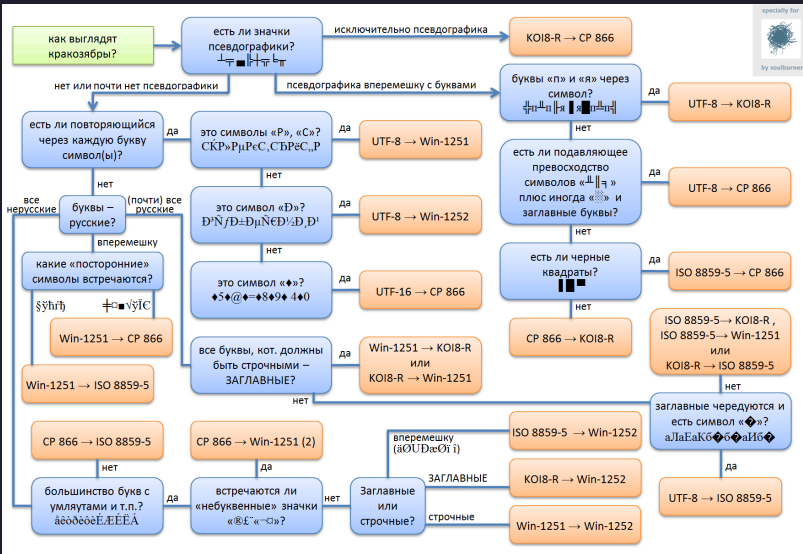
ОПХБЕР ЛХП!

```
$ echo "оПХБЕР ЛХП!" \
    | iconv -t cp1251 \
    | iconv -f koi8-r
Норддн куд!
```

```
$ echo "оПХБЕР ЛХП!" \
  | iconv -t koi8-r \
  | iconv -f cp1251
Привет мир!
```

## Инструменты

- iconv
- enca
- uchardet



# Unicode

## Unicode Standard

- Универсальное кодирование символов
- Количество различных кодов 1 114 112
- Unicode 16.0 (2024 год) использует лишь 154 998

## Цели

- **Универсальность** Содержит все возможные символы современных и древних языков, технических текстов, диакритику и emoji
- **Эффективность** *Plain text* кодирование в одном из трех стандартных форматов: UTF-32, UTF-16, UTF-8
- **Однозначность** Каждый код *однозначно* соответствует единственному символу



Q&A