

NAME

tcs translate character sets

SYNOPSIS

tcs [-slcv] [-f *ics*] [-t *ocs*] [*file ...*]

DESCRIPTION

Tcs interprets the named *file(s)* (standard input default) as a stream of characters from the *ics* character set or format, converts them to runes, and then converts them into a stream of characters from the *ocs* character set or format on the standard output. The default value for *ics* and *ocs* is *utf*, the UTF encoding described in *utf* (6). The *-l* option lists the character sets known to *tcs*. Processing continues in the face of conversion errors (the *-s* option prevents reporting of these errors) The *-c* option forces the output to contain only correctly converted characters; otherwise, 0x80 characters will be substituted for UTF encoding errors and 0xFFFF characters will be substituted for unknown characters.

The *-v* option generates various diagnostic and summary information on standard error or makes the *-l* output more verbose.

Tcs recognizes an ever-changing list of character sets. In particular, it supports a variety of Russian and Japanese encodings. Some of the supported encodings are

<i>utf</i>	The Plan9 UTF encoding, known by ISO as UTF-8
<i>utf1</i>	The deprecated original UTF encoding from ISO 10646
<i>ascii</i>	7-bit ASCII
8859-1	Latin-1 (Central European)
8859-2	Latin-2 (Czech..Slovak)
8859-3	Latin-3 (Dutch..Turkish)
8859-4	Latin-4 (Scandinavian)
8859-5	Part5 (Cyrillic)
8859-6	Part6 (Arabic)
8859-7	Part7 (Greek)
8859-8	Part8 (Hebrew)
8859-9	Latin-5 (Finnish.Portuguese)
<i>koi8</i>	KOI-8 (GOST 19769-74)
<i>jis-kanji</i>	ISO 2022-JP
<i>ujis</i>	EUC-JX:JIS 208
<i>ms-kanji</i>	Microsoft or Shift-JIS
<i>jis</i>	(from only) guesses between ISO 2022-JP, EUC or Shift-Jis
<i>gb</i>	Chinese national standard (GB 2312-80)
<i>big5</i>	Big5 (HKU version)
<i>unicode</i>	Unicode Standard 1.0
<i>tis</i>	Thai character set plus ASCII (TIS 620-1986)
<i>msdos</i>	IBM PC: CP 437
<i>atari</i>	Atari-ST character set

EXAMPLES

```
tcs -f 8859-1
    Convert 8859-1 (Latin-1) character set to UTF format.

tcs -s -f jis
    Convert character encoded in one of several shift-JIS encodings into UTF format. Unknown
    Kanji will be converted into 0xFFFF characters.

tcs -lv
    Print an up to date list of the supported character sets.
```

SOURCE

/sys/src/cmd/tcs

SEE ALSO

ascii (1) *rune* (2) *utf* (6).