

**NAME**

nkf – ネットワーク用漢字コード変換フィルタ v2.0.7

**SYNOPSIS**

```
nkf B<[-butjnesliohrTVvwWJESZxXFfmMBOcdILg]> B<[>I<file ...>B<]>
```

**DESCRIPTION**

**nkf** はネットワークでメールやニュースの読み書きをするために作られた、漢字コードの変換フィルタである。

この**nkf**の特徴としては、入力漢字コード系の統計的な自動認識機能がある。このため、利用者は、入力漢字コード系が何であるかを知らなくても、出力漢字コード系のみ指定すれば良いことになる。ただ、この判定機構は、理論的には完全ではないが、通常のニュースやメールのメッセージについては確実に動作する安全なものにはなっている。

現在、**nkf**が認識できる入力の漢字コード系は、いわゆる JIS コード (ISO-2022-JP に基づくもの)、Shift\_JIS (MS 漢字コード)、日本語 EUC (AT&T コード)、UTF-8、UTF-16 のいずれかである。出力する漢字コード系も、これらである。入力は、ファイルを指定しなければ、標準入力となる。出力は標準出力である。

**OPTIONS**

指定できるオプションは、以下の通り。-**mu** のように続けることができる。

-**j** JIS コードを出力する。(デフォルト)

-**e** EUC コードを出力する。

-**s** Shift\_JIS コードを出力する。

-**w -w8[0] -w16[BL][0]**  
Unicode を出力する。

-**w -w80**  
UTF8 コードを出力する。(BOM 無し)

-**w8**  
UTF8 コードを出力する。

-**w16 -w16B0**  
UTF16 コードを出力する。(Big Endian / BOM 無し)

-**w16B**  
UTF16 コードを出力する。(Big Endian / BOM 有り)

-**w16L**  
UTF16 コードを出力する。(Little Endian / BOM 有り)

-**w16L0**  
UTF16 コードを出力する。(Little Endian / BOM 無し)

-**m[BQSN0]**  
MIME を解読する。(デフォルト) ISO-2022-JP (B encode) と ISO-8859-1 (Q encode) のみを解読する。ISO-8859-1 (Latin-1) を解読する時は、-**I** フラグも必要である。-**m0** では MIME を解読しない。-**mB**、-**mQ** では、BASE64、Q encode されているものとして処理する。

-**mB**  
MIME base64 stream を解読する。ヘッダなどは取り除くこと。

-**mQ**  
MIME quoted stream を解読する。

-**mS**  
MIME のチェックを厳しくする (デフォルト)

- mN**  
MIME のチェックを緩くする
- m0**  
MIME を解読しない。
- mS** は、宣言の内容と、その後続く `encoded-text` の整合性をチェックする。 **-mN** は、改行で切られた MIME など解読する。解読が止まらないこともある。
- M[BQ]**  
MIME に変換する。 JIS にコード変換してから BASE64 される。
- M**  
ヘッダ形式に変換する
- MB**  
base64 stream に変換する
- MQ**  
Quoted stream に変換する
- J -E -S -W**  
期待される入力コードの性質を指定する。
  - J** ISO-2022-JP を仮定する。
  - E** 日本語 EUC (AT&T) を仮定する。
  - S** Shift\_JIS を仮定する。いわゆる半角カナ (JIS X 0201 片仮名) も受け入れる。
  - W**  
UTF-8 を仮定する。
  - W8**  
UTF-8 を仮定する。
  - W16**  
UTF-16 (Little Endian) を仮定する。
  - W16B**  
UTF-16 (Big Endian) を仮定する。
  - W16L**  
UTF-16 (Little Endian) を仮定する。
- x** 通常おこなわれる、いわゆる半角カナ (JIS X 0201 片仮名) からいわゆる全角カナ (JIS X 0208 片仮名) への変換を行わず、半角カナを保存する。入力は、Shift\_JIS の 1byte カナ、SO/SI、ESC-(**-I**, SSO を受け付ける。出力は、日本語 EUC 中では SSO、JIS コード中では ESC-(**-I** をデフォルトで用いる。
- X** いわゆる半角カナ (JIS X 0201 片仮名) をいわゆる全角カナ (JIS X 0208 片仮名) へと変換する。
- B** 壊れた (Broken) JIS コード。ESC がなくなったと仮定する。
  - B1**  
ESC-(, ESC-\$ のあとのコードを問わない
  - B2**  
改行のあとに強制的に ASCII に戻す
- f[m[-n]]**  
一行  $m$  文字になるように、マージンを  $n$  として簡単な整形をおこなう。デフォルトは文字数は 60、マージンは 10 である。
- Z[0-3]**  
JIS X 0208 英数字と若干の記号を ASCII に変換する。つまり、全角を半角に変換する。

- Z -Z0**  
Convert X0208 alphabet to ASCII.
- Z1**  
JIS X 0208 和字間隔を ASCII space 一つに変換する。
- Z2**  
JIS X 0208 和字間隔を ASCII space 二つに変換する。
- Z3**  
>、<、、&、を &gt;、&lt;、&quot;、&amp; に変換する。
- b** バッファリング出力を行う。(デフォルト)
- u** 出力時に、バッファリングしない。 ssh localhost | nkf -u というように使う。
- t** 何もしない。
- I** ISO-2022-JP 以外の漢字コードを = に変換。
- i[@B]**  
JIS 漢字を指示するシーケンスを指定する。(デフォルトは、ESC-\$(B))
- o[BJH]**  
1 バイト英数文字セットを指示するシーケンスを指定する。(デフォルトは、ESC-(-B))
- r** ROT13/47 の変換をする。
- g** 自動判別の結果を出力します。
- v** バージョンを表示する。
- T** テキストモードで出力する。(MS-DOS 上でのみ効力を持つ)
- I** 0x80-0xfe のコードを ISO-8859-1 (Latin-1) として扱う。 JIS コードアウトプットとの組合せのみ有効。 **-s**、**-e**、**-x** とは両立しない。
- O** ファイルに出力する。 UNIX では不要な機能だが Windows や MSDOS では必要らしい。直後の引き数でなく、最後のファイル名が出力ファイル名となり上書きされてしまうので注意。ファイル名がない場合は nkf.out。
- L[uwm] -d -c**  
改行コードを変換する。
  - Lu -d**  
unix (LF)
  - Lw -c**  
windows (CRLF)
  - Lm**  
mac (CR)
 デフォルトでは変換しない。
- ic=<input\_codeset --oc=<output\_codeset>>**  
入力・出力の漢字コード系を指定します。
- ISO-2022-JP**  
いわゆる JIS コード。 **-j**、**-J** と同じ。
- ISO-2022-JP-1**  
RFC 2237 に定められた形式。 JIS X 0212 を含む。
- ISO-2022-JP-3**  
RFC 2237 に定められた形式。 JIS X 0213 を含む。

**EUC-JP**

EUC コード。 **-e**, **-E** と同じ。

**EUC-JISX0213**

文字集合に JIS X 0213:2000 を用いた EUC-JP。

**EUC-JIS-2004**

文字集合に JIS X 0213:2004 を用いた EUC-JP。

**eucJP-ascii**

オープングループ日本ベンダ協議会が定義した eucJP-ascii。 **-x** が暗黙のうちに指定される。

**eucJP-ms**

オープングループ日本ベンダ協議会が定義した euc-JPms。 **-x** が暗黙のうちに指定される。

**CP51932**

Microsft Code Page 51932。 **-x** が暗黙のうちに指定される。

**Shift\_JIS**

Shift\_JIS。 **-s**, **-S** と同じ。

**Shift\_JISX0213**

文字集合に JIS X 0213:2000 を用いた Shift\_JIS。

**Shift\_JIS-2004**

文字集合に JIS X 0213:2004 を用いた Shift\_JIS。

**CP932**

Microsft Code Page 932。 **-x** が暗黙のうちに指定される。

**UTF-8 UTF-8N**

BOM 無しの UTF-8。 **-w**, **-W** と同じ。

**UTF-8-BOM**

BOM 付きの UTF-8。 **-w8** または **-W** と同じ。

**UTF8-MAC**

UTF8-MAC。互換分解されたひらがな・カタカナ等を結合します。

**UTF-16 UTF-16BE-BOM**

BOM 有りて Big Endian の UTF-16。 **-w16B**, **-W16B** と同じ。

**UTF-16BE**

BOM 無しで Big Endian の UTF-16。 **-w16B0**, **-W16B** と同じ。

**UTF-16LE-BOM**

BOM 有りて Little Endian の UTF-16。 **-w16L**, **-W16L** と同じ。

**UTF-16LE**

BOM 無しで Little Endian の UTF-16。 **-w16L0**, **-W16L** と同じ。

**--fj --unix --mac --msdos --windows**

これらのシステムに適した変換をします。

**--jis --euc --sjis --mime --base64**

対応する変換をします。

**--hirakana --katakana**

平仮名、片仮名変換

**--fb-{skip, html, xml, perl, java, subchar}**

Unicode から Shift\_JIS, EUC-JP, ISO-2022-JP に変換する際に、変換できなかった文字をどう扱うかを指定できます。

**--prefix=escape character target character..**

EUC-JP から Shift\_JIS への変換の際、2 バイト目に現れた文字の前にエスケープ文字をつけることができます。引数の 1 文字目がエスケープ文字、2 文字目以降にエスケープされるべき文字を指定します。

例えば、

--prefix=\<\$@ とすると、Shift\_JIS の 2 文字目に \$ か @ が来たら、その前に \ が挿入されます  
 --prefix=@@ とすると、Shift\_JIS の 2 文字目に @ が来たら、その前に @ が挿入されます

**--no-cp932ext**

CP932 において拡張された、NEC 特殊文字、NEC 選定 IBM 拡張文字 (89-92 区)、IBM 拡張文字を変換しません。

**--no-best-fit-chars**

Unicode からの変換の際に、往復安全性が確保されない文字の変換を行いません。Unicode から Unicode の変換の際に -x と共に指定すると、nkf を UTF 形式の変換に用いることができます。(逆に言えば、これを指定しないと一部の文字が保存されません)

パスに関わる文字列を変換する際には、このオプションを指定することを強く推奨します。

**--cap-input, --url-input**

それぞれ :, % に続く 16 進数を文字に変換する

**--numchar-input**

&#...; のような Unicode 文字参照を変換する

**--in-place[=*SUFFIX*] --overwrite[=*SUFFIX*]** 元のファイルを変換結果で置き換える。複数のファイルを書き換えることも可能。元のファイルのタイムスタンプとパーミッションが保持される。現在、作成日時や inode は変更されるが、将来にわたってこの実装のままである保証は無い。

**--guess**

自動判別の結果を出力する

**--help**

コマンドの簡単な説明を表示する。

**--version**

nkf のバージョンを表示する。

-- これ以降のオプションを無視する

**--exec-in** nkf [options] **--exec-in** cmd args... とやると、cmd の出力を nkf の入力とする (config.h で EXEC\_IO を define してコンパイルした時のみ有効)

**--exec-out**

nkf [options] **--exec-out** cmd args... とやると、nkf の出力を cmd の入力とする (config.h で EXEC\_IO を define してコンパイルした時のみ有効)

**BUGS**

実は、入力のコード系の判定は、Shift\_JIS と EUC との自動判定であり、最初に特定できる文字が来た時点で確定してしまう。そして、特定不能の間は保留バッファにためておかれ、確定後に処理される。このため、7 ビット JIS は常に認識される。Shift\_JIS、EUC 混合もほとんどの場合は問題がない。ただし、Shift\_JIS のいわゆる半角カナ (JIS X 0201 片仮名) と EUC 漢字が識別できない。したがって、デフォルトでは Shift\_JIS のいわゆる半角カナはないと仮定している。(つまり、Shift\_JIS か EUC か迷った時は EUC とみなす)

Unicode の日本語以外の部分の変換には対応していないが、**--no-best-fit-chars** を指定すれば、UTF-8 と UTF-16 の相互変換をすることは可能である。

**ACKNOWLEDGE**

以下のかたがた、どうもありがとうございました。

**nkf 1.4**

morb@fujitsu, kiss@ayumi.stars.flab.fujitsu, cen122@flab.fujitsu, yuki@flab.fujitsu 他、fujitsu & flab.fujitsu の皆さんの協力に感謝。 shinoda@cs.titech, kato@cs.titech, uematsu@cs.titech TNX kono@ie.u-ryukyu.ac.jp

**nkf 1.9-**

河野真治  
Akio Furukawa  
Andy Taki  
From: OHARA Shigeki  
Fumitaka Kitagawa  
Hiroaki Sengoku  
Ikuhiro MORITA (森田 育宏)  
Yoshiharu ITO  
Junn Ohta  
KAWAMURA Masao  
Kazuhiko Mori  
Keitaro Isokawa  
Ken-ichi Hirose  
Ki-ichiro SATO  
Kiwamu Aoyama  
Koichi Hirayama  
Mitsuru Hase (長谷 満)  
OHARA Shigeki (大原 重樹)  
Rei FURUKAWA  
Satoru Takabayashi  
Shigeyuki Takagi  
Shin MICHIMUKO  
Tadamasa Teranishi  
TOYODA Jiro  
TSUCHIYA Masatoshi  
Tutomu Sakai  
YAMASHITA Junji (山下 純司)  
Yasuyuki Sato  
Yoshiaki Yanagihara  
hat@so-net  
DCC技術部渡辺  
かとべ / 加藤 貴司  
かべdais  
ひろせ まさあき  
イントラネットシステム(株)松尾  
鷓飼文敏  
塩崎 毅彦(SHIOZAKI Takehiko)  
河村雅夫 (Media Lab.)  
河野 康司  
喜瀬冬猫浩@南国沖縄  
金井 智彦  
桂田 祐史 (かつらだ まさし)  
高橋宜盟  
国吉  
黒江明彦

三宅正泰  
山倉 真  
山本 芳人  
山野裕司 (やまのゆうじ)  
重村法克  
小島基靖  
上田 健  
新井 康司 (Koji Arai)  
森 千絵子  
森 和彦  
水野 貴文  
前地和俊  
仲宗根@並列信頼研  
猪谷 英一郎 == Eiichiro Itani  
藤原 秀行  
八田 真行 (Masayuki Hatta)  
尾川敏也  
望月 孝志  
木村 (@筑波大学数学研究科)  
野村俊彰  
有岡@ミノルタ  
有沢 明宏  
落合一喜  
林 亮  
國安 治  
木下 雄介  
田中隆裕  
鶴谷直樹  
中田伸悦  
わたなべひろふみ  
MoonWolf  
sava

**AUTHOR**

市川 至 ichikawa@flab.fujitsu.co.jp (was ichikawa@fujitsu.JUNET)  
河野 真治 kono@ie.u-ryukyu.ac.jp  
Rei FURUKAWA furukawa@tcp-ip.or.jp  
成瀬  
mastodon

<http://sourceforge.jp/projects/nkf/>

Copyright (C) 1987, FUJITSU LTD. (I.Ichikawa),2000 S. Kono, COW Copyright (C) 2002-2006 Kono, Furukawa, Naruse, mastodon