NAME

KAKASI – Kanji kana simple inverter (between Kanji, both Kana and Romaji)

SYNOPSIS

kakasi [options] [jisyo1 [jisyo2 [jisyo1,,]]]

DESCRIPTION

KAKASI In Japanese sentences are often made up a mixture of Chinese characters (Kanji), Kana (Hiragana and Katakana) and Romaji (Latin phonetical pronunciation). This program converts between these four different ways of writing Japanese.

This program is useful for those whose terminal or desktop does not support the native display of Japanese. Also this is a great tool for those who are learning Japanese (international students and children etc).

A word can be passed into the standard input (stdin), then it is translated and output to standard out (stdout). In the following example the "bunchu" Kanji is converted into Hiragana.

         kakasi –JH < document

Since version 2.3.0 text with spaces in–between words has been supported. In the following example the output has spaces in–between each word.

         kakasi –w < document

Since version 2.3.5 level conversion mode has been supported. In the following example, simple Kanjis are left them unconverted, and difficult Kanjis are translated into Hiragana.

         kakasi –l4 < document

KAKASI It is possible to convert letters to alphabetical characters. Also Katakana letters in the JIS x0201 character set and the Hiragana in the JIS x0208 character set can be converted between each other.

KAKASI The following character set in brackets which is displayed.

ASCII (a)   Known as "ascii" character set.

JISROMAN (j)
         Known as "jis roman" character set.

GRAPHIC (g)
         It is the DEC graphic character set.

Katakana (k)
         JIS x0201, defined as part of the GR character set.

         As a matter of convinience, JIS x0208 is divided as stated below.

Kanji (J)
         JIS x0208 characters included between 16 and 94 sections.

Hiragana (H)
         JIS x0208 characters included in section 4 (Hiragana)

Katakana (K)
         JIS x0208 characters included in section 5 (Katakana)

Sign (E)
         JIS x0208 characters included in section 1,2,3,6,7, and 8. (Note that section 9–15 are undefined in JIS x0208.)

Translation between the following character sets are available.

ASCII   -> JISROMAN, Sign

JISROMAN   -> ASCII, Sign

GRAPHIC   -> ASCII, JISROMAN, Sign

JISx0201 Katakana
                    -> ASCII, JISROMAN, Kana, Hiragana

Sign   -> ASCII, JISROMAN

Katakana   -> ASCII, JISROMAN, JISx0201 Katakana, Hiragana

Hiragana   -> ASCII, JISROMAN, JISx0201 Katakana, Kana

Kanji   -> ASCII, JISROMAN, JISx0201 Katakana, Kana, Hiragana

With conversion of ASCII and the JISROMAN the alphabetical character conversion is done from JISx0201 Katakana, Katakana, Hiragana and Kanji.

Example:

1. All kanji characters are converted to Hiragana.

    kakasi −JH

2. All JIS x0208 characters are converted to JIS X 0201.

    kakasi −Hk −Kk −Jk −Ea

3. All characters are converted to JIS X 0208.

    kakasi −aE −jE −gE −kK

4. All characters are converted to ascii and words are separated.

    kakasi −Ha −Ka −Ja −Ea −ka

5. Exchange between Katakana and Hiragana characters.

    kakasi −HK −KH

CONVERSION DESIGNATED CHARACTER SET

Some character sets are categorized by kakasi and indicated by following mnemonics: a, j, g, k, E, H, K, J.

a −−− ASCII characters
j −−− JIS ROMAN (nearly equal to ASCII, "˜" and "
    different) defined by JIS x0201
g −−− DEC Graphic Characters
k −−− KATAKANA defined by JIS x0201

E, H, K, and J are included in JIS x0208 character set.

J −−− KANJI characters of JIS x0208.
H −−− HIRAGANA characters of JIS x0208.
K −−− KATAKANA characters of JIS x0208.
E −−− Rest of above characters of JIS x0208 which includes
    alphabets, numbers, symbols and so on.

−(from)(to) means conversion from character set (from) to (to). For example, −JK option causes KANJI characters are converted to HIRAGANA. Combinations in the following table are available. (You must not remember it, because the −h shows same information)

```
to\from|  a   j   k   E   H   K   J   g
−−−−−−−+−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−
  a  |  −   o   o1  o   o1  o1  o12 o
  j  |  o   −   o1  o   o1  o1  o12 o
  k  |          −       o   o   o2
  E  |  o   o       −               o
  H  |          o       −   o   o2
  K  |          o       o   −
```

o −− converted.
1 −− converted to Romaji.
2 −− Kanji −> Kana conversion.

KANJI CODING CONVERSION

Unfortunately, several coding systems are used in Japan and JIS x0208 standard are changed at 1983. Therefore, KAKASI can automatically distinguish the coding system and coding revision and then use the same output coding system if the document does not include JIS x0201 KATAKANA. If JIS x0201 KATAKANA is included or you wish to change kanji coding system, you may use the next options.

−i : input coding
−o : output coding

jis −− Widely used on the internet. (Ex: fj, jp, .. newsgroups)
    Derived from ISO−2022 coding manner.
    newjis: JISx0208 (1983) invoked by ESC−$−B.
    oldjis: JISx0208 (1978) invoked by ESC−$−@.
euc,dec −− Often used in UNIX like computers. JISx0208 is
    assigned to GR (MSB is 1). The major difference between
    euc and dec is assignment of JISx0201 KATAKANA and
    the DEC graphic character.
sjis −− Defined by Microsoft Corp. Widely used on the personal
    computers (MSDOS, Mac, ..)
utf8 −− Current international standard. All modern OSs use this
    encoding of the Unicode character set as the default.

ROMAJI CONVERSION

Kanji kana conversion options. Used with −J? option. There are 2 types of Romaji writing. The first is the Kunrei method defined by Japanese government, and the second is the Hepburn method. I think Hepburn method sounds naturally to foreigners.

−rhepburn : Hepburn Method (default)
−rkunrei : Kunrei Method

OTHER OPTIONS

−p: List all possible readings. If there exist two or more
    possible readings, KAKASI shows them in braces {aaa,bbb}.
−s: Insert a separate character between words.
−f: Furigana mode. Shows the original kanji word with reading.
−c: Skip characters within word. (default TAB CR LF BLANK)

−C: Capitalize Romaji word (with −Ja or −Jj option)
−U: Upcase romaji word (with −Ja or −Jj option)
−u: Call fflush().
−w: wakatigaki mode. 'wakatigaki' is word segmentation for
   Japanese sentences.

DICTIONARIES

KAKASI can accept additional dictionary to the system dictionary. The acceptable format of additional dictionary is *SKK* format, and *Wnn* format, and so on. Namely, each record is one line with two fields, Yomi (reading) and Jukugo(idiom). Fields are separated with commas (or TAB, or blank). The kanji code is restricted to JIS or EUC. See another document named JISYO for more details.

ENVIRONMENT VARIABLES

The behavior is affected by the following environment variables.

KANWADICTPATH
   Specifies a path of kanwadict (full−path including filename). Default value is $prefix/share/kakasi/kanwadict.

ITAIJIDICTPATH
   Specifies a path of itaijidict (full−path including filename). Default value is $prefix/share/kakasi/itaijidict.

AUTHOR

Hironobu Takahasi <takahasi@tiny.or.jp>

FILES

$prefix/share/kakasi/kanwadict
   It is a binary dictionary of *KAKASI*. It is automatically converted from kakasidict by mkkanwa when the package is installed.

SEE ALSO

mkkanwa(1)

DIAGNOSTICS

Return status except 0 when there is any trouble.

BUGS

Report bugs to KAKASI Project <kakasi−dev@namazu.org>. Please DO NOT CONTACT to the originator (Takahasi−san).

NOTE ABOUT ENGLISH MANUAL

The content of English manual is not exactly same as that of Japanese manual.