

sjisEUC ユーザーズガイド

sjiseuc コマンドは、シフト JIS のテキストファイルを変換して EUC-JP のテキストファイルに変換します。

下線で代用

EUC-JP は、Windows ではなく UNIX の日本語です。sjiseuc コマンドがシフト JIS から EUC-JP にテキストファイルを変換するとき、たとえば、丸付きの数字 (①②③⑩⑳)、ローマ数字 (ⅠⅡⅢⅣⅩ)、1 バイトのカタカナは、EUC-JP で表現しにくいいため、下線に変更されます。

変換表を編集

たとえば、①、②、③が、それぞれ 1、2、3 で代用されるように、添付の diyocommand は、変換表を編集できます。

改行

テキストファイルに関して、Windows では、0Dh および 0Ah で改行しますが、UNIX では、0Ah のみで改行します。sjiseuc コマンドは、各行から 0Dh を削除できます。

標準入力

ファイル名が指定されていない場合、sjiseuc コマンドは、標準入力から行を読み込むことができます。

インストール

sjiseuc コマンドは、32 ビット版 Windows のコンソールアプリケーションです。

パスを通す手順の例

1. トリシーカー (<http://tori.tobihiro.jp/PDF-zip-7z.html>) というウェブサイトから sjisEUCjpCxx.zip をダウンロードしてください。Cxx は、C++ を意味します。
2. sjisEUCjpCxx.zip を展開してください。
3. C ドライブに sjisEUCjpCxx フォルダを移動してください。
4. sjisEUCjpCxx フォルダの名前を変更しないでください。
5. Windows ロゴキー (田キー) を押しながら Pause を押してください。
6. メニューからシステムの詳細設定を選択してください。
7. 詳細設定タブをクリックしてください。
8. [環境変数] ボタンを押してください。キーワード: システム、詳細、環境変数
9. システム環境変数として、Path が見えるまでスクロールしてください。

10. Path をクリックして選択してください。
11. [編集] ボタンを押してください。
12. Windows 10 の場合、[テキストの編集] ボタンを押してください。
13. 右向き矢印(→)キーまたは End キーを押してください。カーソルが末尾に移動します。
14. フォルダのことをディレクトリとも言います。セミicolon(;)は、ディレクトリどうしを区切る記号です。 ;C:¥sjisEUCjpCxx を入力してください。
15. 各ウィンドウで OK ボタンを押してください。

フォルダごと削除

アンインストール(プログラムの削除)を行うには、sjisEUCjpCxx フォルダごと削除してください。Path から sjisEUCjpCxx を削除してください。

著作権

Sogaya(そがや)は、sjisEUC の著作権を保有しています。ソースコードを添付しました。

免責条項

sjiseuc コマンドのご利用によって発生するいかなる損害も、Sogaya は、責任を負わないものとします。

構築

Dev-C++の IDE において、sjisEUC は、GCC で構築されています。

gcc (GCC) 3.4.2 (mingw-special)

Borland 無料コマンドラインコンパイラ

Borland 無料コマンドラインコンパイラで構築できます。makefile.mak を添付しました。

Borland C++ 5.5.1 for Win32

ユーザーズガイド

2019 年 6 月 9 日 Revision 1.02909

sjiseuc コマンドのパラメータ

MS932~EUC-JP.bin というバイナリファイルにしたがって、sjiseuc コマンドは、シフト JIS

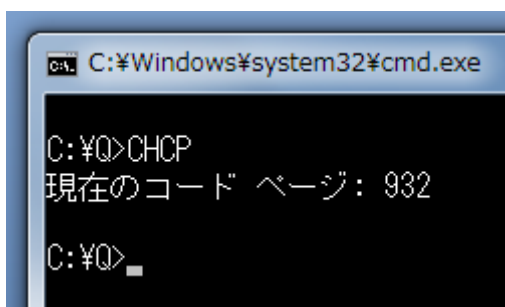
のテキストファイルを EUC-JP のテキストファイルに変換します。

コマンドプロンプト

Windows のコマンドプロンプトは、シフト JIS が多用されます。MS932 は、いわゆるシフト JIS ですが、規格外です。たとえば、丸付きの数字(①②③⑩⑰)、ローマ数字(ⅠⅡⅢⅣⅤ)、1 バイトのカタカナは、EUC-JP で表現しにくい。

CHCP コマンド

コマンドプロンプトの文字コードがシフト JIS に設定されていることを確認するには、パラメータを付加しないで、コマンドプロンプトに CHCP を入力して、Enter を押してください。932 が表示されます。MS932 が使用されることがわかります。どの Windows も最初から CHCP コマンドを用意しています。MS932 は、[CP932](#)とも言います。



```
C:\Windows\system32\cmd.exe
C:\>CHCP
現在のコード ページ: 932
C:\>
```

MORE コマンド

コマンドプロンプトでシフト JIS のテキストファイルを閲覧する方法として、MORE コマンドが利用できます。スペースキーで進行します。Q で終了します。どの Windows も最初から MORE コマンドを用意しています。

atnd コマンド

添付の atnd は、シフト JIS 用のコンソールアプリケーションです。シフト JIS のファイルの最初から 7 行を表示するには、コマンドプロンプトに、たとえば、下記のコマンドおよびパラメータを入力して、Enter を押してください。-h の直後に 7 を入力してください。-h7 は、-b6 -p7 を意味します。

```
atnd -h7 Gongitsune.txt
```

```
C:¥Q>atnd -h7 Gongitsune.txt
こん狐
新美南吉
—

これは、私《わたし》が小さいときに、村の茂平《もへい》というおじいさんからきいたお話です。
むかしは、私たちの村のちかくの、中山《なかやま》というところに小さなお城があって、中山さまというおとのさまが、おられたそうです。

C:¥Q>
```

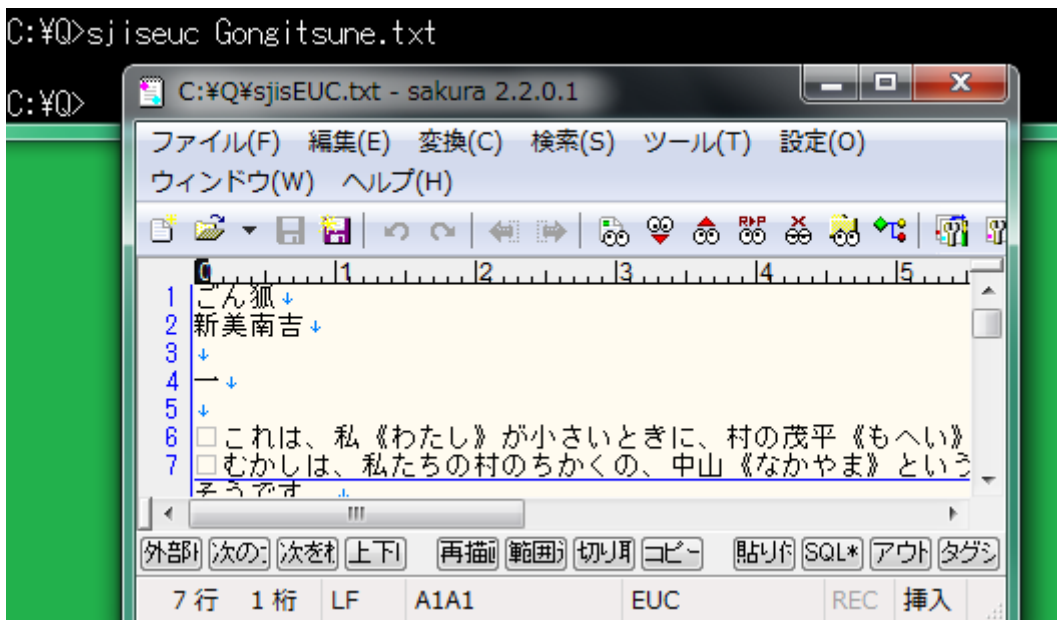
テキストエディタ

EUC-JP をサポートしているテキストエディタとしては、EmEditor, K2Editor, MIFES, TeraPad, WZ, サクラエディタなどがあります。

変換したいファイル

sjiseuc コマンドのパラメータとして、シフト JIS のファイルを指定してください。sjiseuc コマンドは、シフト JIS のファイルを読み込んで、sjisEUC.txt という EUC-JP のファイルを書き出します。

sjiseuc Gongitsune.txt



2 個のファイル

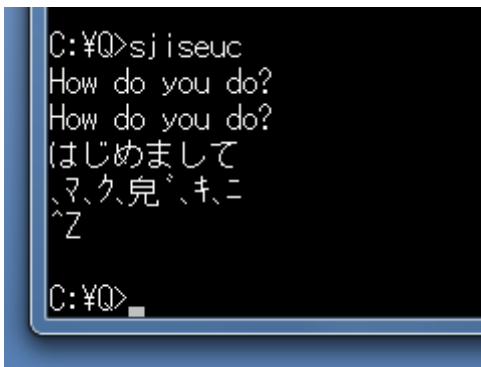
シフト JIS のファイル、EUC-JP のファイルがこの順に指定された場合、sjiseuc コマンドは、シフト JIS のファイルから行を読み込んで、EUC-JP のファイルに行を書き出します。

す。

sjiseuc Shift_JIS.txt EUC-JP.txt

標準入力

sjiseuc コマンドのパラメータとして、ファイル名が指定されなかった場合、sjiseuc コマンドは、標準入力から行を読み込んで、標準出力に書き出します。下記の例では、英語を入力した場合、英語が表示されますが、日本語を入力した場合、文字化けが発生することがわかります。コマンドプロンプトは、シフト JIS を正しく表示します。EUC-JP は、文字化けが発生します。Ctrlを押しながら Z を押すことでファイルの終わり(EOF)を入力しました。^Z が表示されました。



```
C:¥Q>sjiseuc
How do you do?
How do you do?
はじめまして
マ、ク、兒、キ、ニ
^Z
C:¥Q>
```

リダイレクト

標準入力から行を読み込む場合、リダイレクトでファイルを作成できます。ただし、0Dh および 0Ah で改行します。

sjiseuc < Shift_JIS.txt > EUC-JP.txt

0Dh を削除

テキストファイルに関して、Windows では、0Dh および 0Ah で改行しますが、UNIX では、0Ah のみで改行します。標準入力から読み込んで標準出力へ書き出す場合、0Dh が含まれます。さもなければ、sjiseuc コマンドは、各行から 0Dh を削除します。

-CRLF

0Dh および 0Ah で改行するには、パラメータとして、-CRLF を付加してください。

-LF

リダイレクトではなく sjiseuc コマンドでファイルを作成する場合、0Ah のみで改行するために、パラメータとして、-LF を付加できます。

iconv コマンド

文字コードを変換するために、iconv コマンドが GNU Win32 に用意されています。インターネットで、たとえば、libiconv-1.9.2-1 などのインストーラーを検索してください。接頭辞の lib は、ライブラリを意味します。このインストーラーをダウンロードして、実行して、32 ビット版 Windows 7 のパソコンに iconv.exe をインストールしました。

FC コマンド

テキストファイルを比較するために、FC コマンドを試すことができます。どの Windows も最初から FC コマンドを用意しています。iconv コマンドの結果と、sjiseuc コマンドの結果をテキストファイルとして比較しました。各行が一致しました。さもなければ、一致しない付近が表示されます。

```
C:¥Q>iconv -f Shift_JIS -t EUC-JP Gongitsune.txt > text.tmp
C:¥Q>sjiseuc Gongitsune.txt
C:¥Q>FC text.tmp sjisEUC.txt
ファイル text.tmp と SJISEUC.TXT を比較しています
FC: 相違点は検出されませんでした
C:¥Q>
```

COMP コマンド

バイナリファイルとして比較するには、FC コマンドのパラメータとして、/B を付加してください。Windows 2000, Windows XP, Windows Vista, Windows 7, Windows 10 では、COMP コマンドを試すことができます。iconv コマンドの結果と、-CRLF 付きの sjiseuc コマンドの結果をバイナリファイルとして比較しました。0Dh および 0Ah による改行も一致しました。

```
C:¥Q>iconv -f Shift_JIS -t EUC-JP Gongitsune.txt > binary.tmp
C:¥Q>sjiseuc -CRLF Gongitsune.txt
C:¥Q>COMP binary.tmp sjisEUC.txt
binary.tmp と sjisEUC.txt を比較しています...
ファイルに違いはありません
(ほかのファイルを比較しますか (Y/N)? n
C:¥Q>
```

画面を消去

コマンドプロンプトの画面を消去するには、コマンドプロンプトに下記のコマンドを入力して、Enter を押してください。

CLS

バージョン番号

sjiseuc コマンドのバージョン番号を確認するには、コマンドプロンプトに下記のコマンドおよびパラメータを入力して、Enter を押してください。コマンドおよびパラメータの例も表示されます。

sjiseuc -V

```
C:¥Q>sjiseuc -V
sjisEUC C++ Version 1.16

sjisEUC SJIS
sjisEUC SJIS EUC
sjisEUC < SJIS > EUC
sjisEUC -a SJIS
sjisEUC -b SJIS
sjisEUC -b SJIS SJIS
sjisEUC -b -y SJIS
sjisEUC -CRLF SJIS
sjisEUC -CRLF SJIS EUC
sjisEUC -d SJIS
sjisEUC -LF SJIS
sjisEUC -s SJIS
sjisEUC -t SJIS EUC
sjisEUC -w SJIS
sjisEUC -V
```

短い説明

コマンドおよびパラメータの各行に短い説明を付与して表示するには、コマンドプロンプトに下記のコマンドおよびパラメータを入力して、Enter を押してください。

sjiseuc -VC

パラメータ	密接に関係するファイル	説明
-a	MS932~EUC-JP.bin	sete.txt から変換表のパス名を取得します。
-b	binary.tmp	いったん binary.tmp を作成して、改名します。
-CRLF	EUC-JP テキスト	0Dh および 0Ah で改行します。
-d	MS932~EUC-JP.bin	変換表を探索してはならない
-LF	EUC-JP テキスト	0Ah のみで改行します。
-s	シフト JIS テキスト	0Ah 以外の制御文字を削除します。
-t	シフト JIS テキスト、EUC-JP テキスト	シフト JIS のテキストは、EUC-JP のテキストと同じ日時を有します。
-w	MS932~EUC-JP.bin	どの変換表を使用するか表示します。
-V		バージョン番号および用法を表示します。
-VC		用法、短い説明を表示します。
-y	binary.tmp	binary.tmp を削除してはならない。

実行ファイル

変換表の名前は、MS932~EUC-JP.bin です。sjiseuc コマンドの特徴として、実行ファイル (sjisEUC.exe) の外部から変換表を読み込むことができます。実行ファイルは、自分が**インストールされたディレクトリ**に存在している変換表を読み込みます。

カレントディレクトリ

変換表および実行ファイルが同じディレクトリに存在していない場合、カレントディレクトリから変換表を読み込みます。

Path に記載の各ディレクトリ

変換表を開くことができるまで、sjiseuc コマンドは、下記の順に変換表を探索します。

1. インストールされたディレクトリ
2. カレントディレクトリ
3. 環境変数 (Path) に記載の各ディレクトリ
4. 利用者が設定ファイル (sete.txt) に記入したパス名

探索してはならない

変換表の探索を抑制するには、sjiseuc コマンドに -d を付加してください。

設定ファイルを優先

設定ファイル(sete.txt)に記入の変換表を優先して使用するには、sjiseuc コマンドのパラメータとして、-a を付加してください。

どの変換表を使用するか

過去に実行ファイルがインストールされたディレクトリか、Path にあるディレクトリか、カレントディレクトリのいずれかに存在している変換表を開くことができたとき、どの変換表を使用するか表示するには、sjiseuc コマンドのパラメータとして、-w を付加してください。パス名ではなく単なるファイル名が表示された場合、カレントディレクトリの変換表を使用します。

sjiseuc Gongitsune.txt -w

```
C:¥Q>sjiseuc Gongitsune.txt -w
C:¥sjisEUCjpcxx¥MS932~EUC-JP.bin is primarily accessible to sjisEUC
C:¥Q>_
```

新しいファイル

COPY コマンドで、シフト JIS の新しいテキストファイルを作成するには、コピーされるファイルとして CON を指定してください。この場合、CON は、パソコンのキーボードです。コマンドプロンプトは、CON がファイルであるかのように COPY コマンドを実行します。

ファイルの終わり

コマンドプロンプトに、ファイルの終わり(EOF)を入力するには、Ctrl を押しながら、Z を押してください。Ctrl + Z は、この操作を意味します。^Z が画面に表示される場合があります。

中止したい合図

コマンドプロンプトに、中止したい合図を入力するには、Ctrl を押しながら、C を押してください。Ctrl + C は、この操作を意味します。^C が画面に表示される場合があります。

かな漢字変換

コマンドプロンプトで、かな漢字変換を行うには、Alt を押しながら[半角/全角]キーを押してください。Windows 7, Windows 10 の場合、Alt を押す必要がありません。かな

漢字変換モードから脱出するには、[半角／全角]キーを押してください。

F8 で半角カタカナに変換

コマンドプロンプトをかな漢字変換モードに移行して、ローマ字入力法で半角カタカナに変換する手順の例を示します。

1. ECHO を入力してください。スペースキーを押して空白を入力してください。まだ Enter キーを押さないでください。
2. Alt を押しながら[半角／全角]キーを押してください。かな漢字変換モードに移行します。
3. H キーを押してください。
4. E キーを押してください。へが表示されます。
5. F8 を押してください。へからへに変化します。
6. Enter キーを押してください。かな漢字変換が確定します。
7. もう一回 Enter キーを押してください。ECHO コマンドおよびパラメータが入力されます。
8. コマンドプロンプトに半角カタカナのへが表示されます。
9. かな漢字変換モードから脱出するには、[半角／全角]キーを押してください。

変換表を編集する

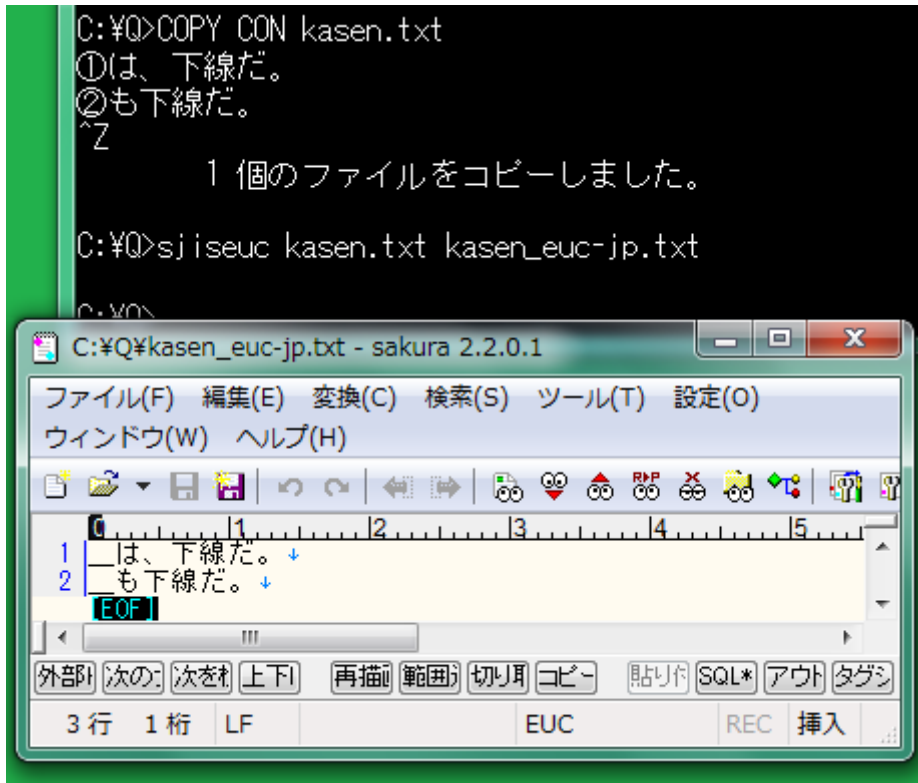
変換表の名前は、MS932~EUC-JP.bin です。通常、変換表および実行ファイルは、同じディレクトリに存在しています。

diyo コマンド

添付の diyo コマンドは、MS932~EUC-JP.bin というバイナリファイルを編集できます。

すべて下線になる

EUC-JP は、Windows ではなく UNIX の日本語です。sjiseuc コマンドがシフト JIS から EUC-JP にテキストファイルを変換するとき、丸付きの数字(①②③⑩⑰)、ローマ数字(I II III IV X)などの機種依存文字は、すべて下線()に変更されます。1 バイトのカタカナは、ASCII の下線()に変更されます。



0Ah のみで改行

EUC-JP のテキストファイルを上に表示しました。0Ah のみで改行しています。サクラエディタのウィンドウの下部にあるステータスバーに LF が表示されます。サクラエディタの改行文字は、左向きではなく下向きの矢印になります。

1 文字ずつ編集

たとえば、①を下線ではなく1で代用するには、コマンドプロンプトに下記のコマンドおよびパラメータを入力して、Enter を押してください。通常の文字および代用文字のシフト JIS 文字コードが十六進数で表示されます。

diyo -s① -e 1

どの変換表を編集するか

diyo コマンドがアクセスできる変換表のパス名を表示するには、コマンドのパラメータとして、-w を付加してください。

diyo -s① -e 1 -w

```

C:¥Q>diyo -s① -e1 -w
C:¥sjisEUCjpCxx¥MS932~EUC-JP.bin is primarily accessible to diyo.
8740 ⇒ 8250

C:¥Q>DIR | FINDSTR "MS932~EUC-JP.bin"
2019/05/13  20:48                31,528 MS932~EUC-JP.bin

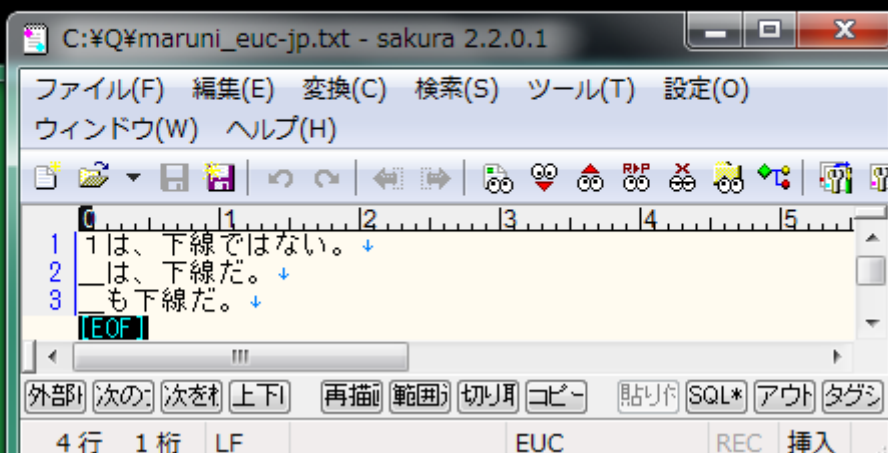
C:¥Q>COPY MS932~EUC-JP.bin C:¥sjisEUCjpCxx /Y
      1 個のファイルをコピーしました。

C:¥Q>COPY CON maruni.txt
①は、下線ではない。
②は、下線だ。
③も下線だ。
^Z
      1 個のファイルをコピーしました。

C:¥Q>sjiseuc maruni.txt maruni_euc-jp.txt

C:¥Q>

```



1 バイトの文字

1 バイトの文字としては、ASCII および半角カタカナがあります。

!	~	#	\$	%	&	'	()	*	+	,	-	.	/	
0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
@	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	[¥]	^	_
`	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	{		}	~	.
。	「	」	、	ヲ	ア	イ	ウ	エ	オ	ヤ	ユ	ヨ	ッ	ッ	ッ
ー	ア	イ	ウ	エ	オ	カ	キ	ク	ケ	コ	サ	シ	ス	セ	ソ
タ	チ	ツ	テ	ト	ナ	ニ	ヌ	ネ	ノ	ハ	ヒ	フ	ヘ	ホ	マ
ミ	ム	メ	モ	ヤ	ユ	ヨ	ラ	リ	ル	レ	ロ	ワ	ヰ	ヱ	。

1 バイトのカタカナ

1 バイトのカタカナは、シフト JIS に含まれています。コマンドプロンプトの表示では、1

バイトのカタカナは、幅が漢字の半分になります。

半角カタカナ

1 バイトのカタカナを半角カタカナとも言います。半角カタカナの代用文字として、ASCII の下線 (_) が sjiseuc コマンドに使用されます。sjiseuc コマンドが、たとえば、ナ の代用として ASCII の下線ではなく N を使用するように変換表を編集するには、コマンドプロンプトに下記のコマンドおよびパラメータを入力して、Enter を押してください。k およびナ の間に空白を挿入しないでください。

diyo -k ナ -aN

設定ファイル

diyo コマンドの設定ファイルは、setting.txt です。setting.txt は、必須ではありません。setting.txt を開くことができる場合も、diyo コマンドのパラメータで設定を変更できません。setting.txt による設定を diyo コマンドのパラメータで変更できないと思われる場合、setting.txt を削除してください。

パラメータ	用語	目的	説明
-a	ASCII	半角カタカナを廃止	1 バイトの英字、数字、記号
-e	EUC-JP	機種依存文字を廃止	UNIX の日本語
-k	katakana		半角カタカナ
-s	Shift_JIS		Windows の日本語、MS932

記入した設定を否定するパラメータ

たとえば、-w が記入されていても、diyo コマンドのパラメータとして、-w- を付加すれば、変換表のパス名が表示されません。

記入した設定	否定する入力	説明
-d	-d-	変換表を探索する / 探索しない
-w	-w-	変換表のパス名を表示する / 表示しない

バージョン番号

diyo コマンドのバージョン番号を表示するには、コマンドプロンプトに下記のコマンドおよびパラメータを入力して、Enter を押してください。

diyo -V

短い説明

短い説明を表示するには、コマンドプロンプトに下記のコマンドおよびパラメータを入力して、Enter を押してください。

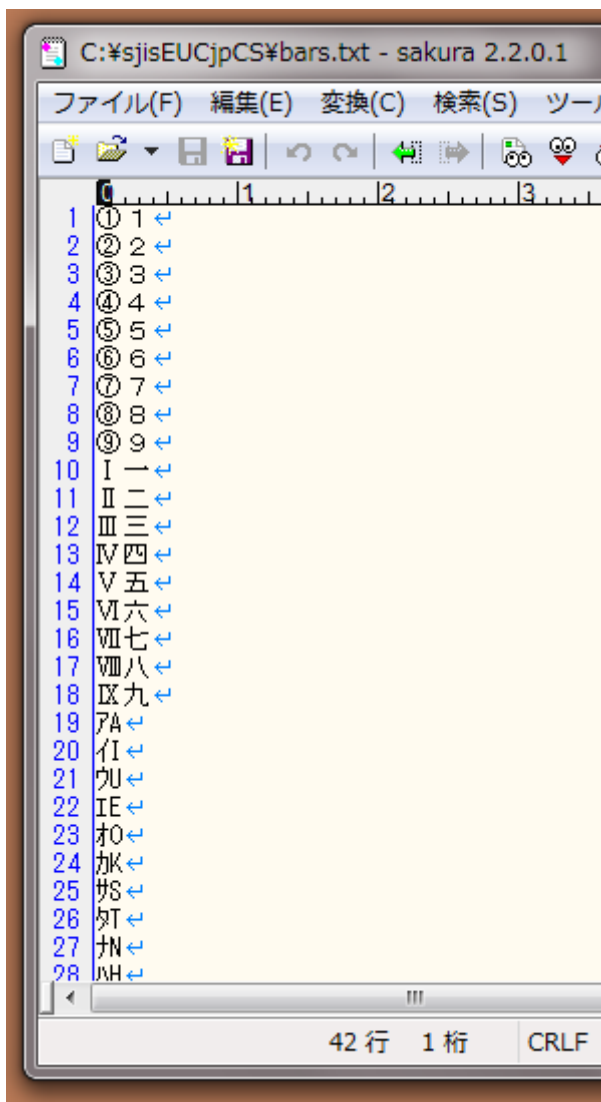
diyo -VC

代用文字のリスト

事前に利用者が代用文字のリストを用意した場合、diyo コマンドは、代用文字のリストにしたがって、一度に多数の文字を編集できます。

通常文字および代用文字

diyo コマンドは、代用文字リストのファイルから 906 行まで読み込むことができます。代用文字リストのファイルの各行は、通常文字と、代用文字とからなります。代用文字もシフト JIS で記入してください。



ファイル名を指定

利用者が代用文字リストのファイルを用意した場合、diyo コマンドのパラメータとして、-f を付加して、空白で区切って、さらにファイル名を付加してください。

diyo -f bars.txt

どの変換表を使用するか

過去に実行ファイルがインストールされたディレクトリか、Path にあるディレクトリか、カレントディレクトリのいずれかに存在している変換表を開くことができたとき、どの変換表を使用するか表示するには、diyo コマンドのパラメータとして、-w を付加してください。パス名ではなく単なるファイル名が表示された場合、カレントディレクトリの変換表を使用します。

diyo -f bars.txt -w

```
C:¥Q>diyo -f bars.txt -w
C:¥s jisEUCjpCxx¥MS932~EUC-JP.bin is primarily accessible to diyo.

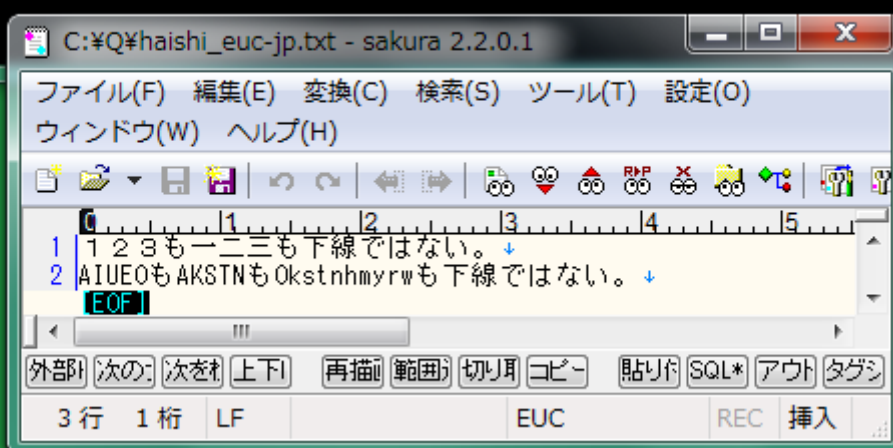
C:¥Q>DIR | FINDSTR "MS932~EUC-JP.bin"
2019/05/13  21:03                31,528 MS932~EUC-JP.bin

C:¥Q>COPY MS932~EUC-JP.bin C:¥s jisEUCjpCxx /Y
      1 個のファイルをコピーしました。

C:¥Q>COPY CON haishi.txt
①②③も I II III も下線ではない。
アイエオも アカクサも オコトホニホヒも下線ではない。
^Z
      1 個のファイルをコピーしました。

C:¥Q>sjiseuc haishi.txt haishi_euc-jp.txt

C:¥Q>
```



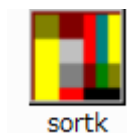
並べ替えに適した代用文字リスト

上記の bars.txt では、ソフトウェアで並べ替えた結果が昇順になりません。代用文字リ

ストのファイルを作成するとき、並べ替えも考慮したい。

sortk コマンド

添付の sortk コマンドは、テキストファイルの行を並べ替えて、標準出力に書き出します。



CSV

拡張子が.csv であるテキストファイルは、各行が、コンマで分離された多数の語句からなります。CSV のファイルは、Excel などの表計算ソフトウェアで容易に表示できます。sortk コマンドは、CSV の欄にある語句を比較して、行を並べ替えます。

区切り文字

CSV の行を並べ替えるには、コマンドプロンプトに sortk コマンドのパラメータとして、`-t` (マイナスティーコンマ) を付加してください。`-t` およびコンマの間に空白を挿入しないでください。

空白

区切り文字が指定されていない場合、sortk コマンドは、区切り文字として空白を使用します。

標準入力

CSVファイル名が指定されていない場合、sortkコマンドは、標準入力から読み込みます。

新しいファイル

下記のとおり、標準出力からファイルにリダイレクトすることで、新しいファイルを作成できます。

sortk `-t, < csv > text.tmp`

命名

csv を読み込んで、新しいファイルの作成および命名を行うには、コマンドプロンプトに sortk コマンドのパラメータとして、`-o text.tmp` を付加してください。`-o` および text.tmp の間に空白を挿入してください。text.tmp は、新しいファイルの名前です。

sortk -t, -o text.tmp csv

第 2 欄を比較

sortk コマンドは、たとえば、第 2 欄にある語句を比較して、CSV のテキストファイルの行を並べ替えることができます。第 2 欄にある語句を比較するには、sortk コマンドのパラメータとして、-k2,2 を付加してください。-k および数字の間に空白を挿入しないでください。

sortk -t, -k2,2 KeyWords.csv

```
C:\¥Q>sortk -t, -k2,2 KeyWords.csv
合図,アズ,
アクセス,アケス,
ASCII,アスキ,
アルファベット,アルファベッ,
アンインストール,アンインストール,
1バイトのカタカナ,イチバイトカタカナ,
インストーラー,インストール,
インストール,インストール,
インターネット,インターネット,
Windows,ウインドウズ,
Windowsロゴキー,ウインドウズロゴキー,
ウェブサイト,ウェブサイト,
英語,エイゴ,
英字,エイジ,
Excel,エクセル,
閲覧,イツラン,
大文字,オオモジ,
改行,カクキョウ,
拡張子,カクサウジ,
下線,カセン,
かな漢字変換,カナカンジヘンカン,
カレントディレクトリ,カレントディレクトリ,
```

sortk の由来

FreeBSD, Linux, Solaris などの UNIX は、sort コマンドを用意しています。sortk は、UNIX の sort コマンドに似ています。sortk(ソーティーケー)という名前は、UNIX の sort -t, -k に由来します。

日本語

sortk コマンドは、半角カタカナの濁点、半濁点を無視できます。Windows のコマンドプロンプト

ロンプトで、EUC-JP のテキストファイルを並べ替えることができます。0Ah のみで改行
できます。

sjiseuc コマンド

sortk コマンドは、sjiseuc コマンドに添付されています。sortk コマンドは、Windows のコ
ンソールアプリケーションです。32 ビット版 Windows で構築しました。

zip を展開

トリシーカー (<http://tori.tobihiro.jp/PDF-zip-7z.html>) というウェブサイトから
sjisEUCjpCxx.zip をダウンロードしてください。Cxx は、C++ を意味します。
sjisEUCjpCxx.zip を展開してください。diyo, sjiseuc, sortk, narfi, rufeol を使用した例

コマンド	文字コード	説明
atnd	シフト JIS	周辺の行を表示する。最長の行の長さを記録する。
dayu	Shift_JIS	変換表の編集およびシフト JIS テキストの規格化を行う。
diyo	EUC-JP	変換表を編集する。
hanzen	シフト JIS	ある欄で、半角カタカナを全角カタカナに置換する。
keyun	バイナリ	binary.tmp を読み込んで、行を並べ替える
narfi	シフト JIS	広い表から細長い表を作成する。ある欄を編集する。
obot	ASCII	ファイルの開始部分のみ 2 桁の十六進数で表示する。
rufeol	シフト JIS	テキストの各行の末尾から不要な文字を除去する。
sortk	シフト JIS	選択された欄を比較して、テキストの行を並べ替える。
sjiseuc	EUC-JP	シフト JIS テキストを EUC-JP テキストに変換する。

免責条項

sortk コマンドのご利用によって発生するいかなる損害も、Sogaya は、責任を負わない
ものとします。

構築

Dev-C++ の IDE において、C++ のプロジェクトとして GCC で sortk を構築しました。

bcc32

Borland 無料コマンドラインコンパイラで構築できます。makefile.mak を添付しました。

sortk コマンドのパラメータ

sortk コマンドは、CSV の欄にある語句、整数、浮動小数点数を比較して、行を並べ替
えます。利用者は、CSV の欄を選択できます。


バージョン番号

sortk コマンドのバージョン番号を確認するには、-V を付加してください。

短い説明

パラメータに関する短い説明を表示するには、sortk コマンドのパラメータとして、-VC または--help を付加してください。画面からテキストファイルに、リダイレクトするには、-s を前置してください。

sortk -s --help > help.txt



```
C:\>sortk -s --help > help.txt
C:\>NOTEPAD help.txt

help - メモ帳
ファイル(F) 編集(E) 書式(O) 表示(V) ヘルプ(H)
sortk C++ Version 1.195

sortk -t, -o new csv # Read csv to produce a new file.
sortk -t, < csv > new # Read csv to produce a new file.
sortk -t, -a4 -h1011 csv # Compare in Field-1, Field-3 and Field-4.
sortk -t, -j2 csv # Compare phrases in Field-2.
sortk -t, -k2,2 csv # Compare phrases in Field-2.
sortk -t, -k2,3 csv # Compare phrases from Field-2 to Field-3.
sortk -t, -k2 csv # Compare phrases from Field-2 to finish.
sortk -t, -k1 -$ csv # Check if all lines share the field count.
sortk -t, -k2,2 -# csv # Add ID to each summary.
sortk -t, -k2,3 -# -! csv # Report ID if any summaries match each other.
sortk -t, -k2,2 -8 csv # UTF-8
sortk -t, -k2,2 -b csv # Skip space in the first part of Field-2.
sortk -t, -k2,2 -c- csv # Skip combining characters.
sortk -t, -k2,2 -e csv # EUC-JP
sortk -t, -k2,2 -f csv # EIGHT equals eight.
sortk -t, -k2,2 -g csv # Compare 0.12 with 0.78 in Field-2.
sortk -t, -k2,2 -i csv # Skip control characters.
sortk -t, -k2,2 -l csv # Show how large the buffer is.
sortk -t, -k2,2 -n csv # Compare 3456 with 9012 in Field-2.
sortk -t, -k2,2 -p csv # P and Q are omitted from ASCII characters.
sortk -t, -k2,2 -q csv # Commas disappear if placed in quotes.
sortk -t, -k2,2 -r csv # Arrange in descending order.
sortk -t, -k2,2 -y csv # Do not remove binary.tmp
sortk -t, -k2,2 -y -% csv # Produce manager.txt
sortk -t, -k2,2 -z256 csv # Use a buffer of 256 bytes.
sortk -t%l -k2,2 file # Wiki table
sortk -t%l -k2,2 -x- csv # Do not check if the file name includes .csv
sortk -t%x7C -k2,2 file # Wiki table
sortk -U131072 -k2,2 csv # The number of lines shall not exceed 131072.
sortk -v%x1F -k2,3 csv # 1Fh connects phrases from Field-2 to Field-3
sortk -w~ -k2,4 csv # ~ is used as a quote.
```

パラメータ	ほぼ同等な入力	説明
-!		要約が合致した場合、ID を表示する。
-\$		どの行も同じ欄数を有しているか点検する。
-#		要約に ID を付与する。
-%		設定ファイルとして、manager.txt を作成する。
-8	-e	0Ah のみで改行する。
-a4	-h1111000000	第 1 欄から第 4 欄まで選択する。
	-k1,4	
-b		比較開始欄の前半にある空白を無視する。
-c-		半角カタカナの濁点、半濁点を無視する。
-e	-8	0Ah のみで改行する。
-f		大文字と小文字を区別しないで比較する。
-g		浮動小数点数を比較する。
-h1011		第 2 欄を除外する。
-i		タブ、改ページなど制御文字を無視する。
-j2	-k2,2	第 2 欄のみ選択する。
-k2		第 2 欄から最終欄まで選択する。
-k2,3	-a3 -h011	第 2 欄から第 3 欄まで選択する。
-l		何バイトのバッファを使用するか表示する。
-n		整数を比較する。
-o new	> new	新しいファイルを命名する。
-p		濁点の代用文字として、P を無視する。
-q		引用符で囲まれた区切り文字を削除する。
-r		降順で並べ替える。
-s		標準エラー出力ではなく標準出力に書き出す。
-t,	.csv の拡張子	区切り文字としてコンマを使用する。
-t¥x7C	-t¥1	Wiki の表の区切り記号として縦線を使用する。
-U131072		行数は、131072 を超えてはならない。
-v¥x1F		選択された語句を 1Fh で連結する。
-V		バージョン番号および用法を表示する。
-VC	--help	パラメータの短い説明を表示する。
-w~	-w¥x7E	~ が引用符として利用される。
-x-		.csv がファイル名に含まれるか点検しない。
-Y		binary.tmp を削除してはならない。
-z256		256 バイトのバッファを使用する。

¥x および十六進数

上記のパラメータのうち、-t, -v, -w は、十六進数で 1 バイト文字を指定できます。たとえば、-w¥x7E は、-w~と同じ結果を生じます。

¥v で大なり記号を意味する

AWK, C#などのプログラミング言語では、¥v は、垂直タブ(0Bh)を意味する場合がありますが、sortk コマンドのパラメータでは、大なり記号(>)を意味します。すなわち、-t¥v は、-t¥x3E と同じ結果を生じます。下表では、-t の例を示します。-v, -w の表も同様になります。

パラメータ	ほぼ同等な入力	1 バイト文字	説明
-t¥A	-t¥x3C	<	小なり記号
-t¥^	-t¥x3C	<	小なり記号
-t¥l	-t¥x7C		縦線
-t¥s	-t¥x20		空白
-t¥t	-t¥x09		タブ、水平タブ
-t¥v	-t¥x3E	>	大なり記号

パラメータからコンマを排除

sortk コマンドのパラメータとして、-t,(マイナスティーコンマ)も-k2,3(マイナスケーニコマサン)もコンマが含まれていますが、コンマの使用を回避できます。

範囲を表現するハイフン

コンマではなくハイフンを使用できます。-k2,3 ではなく-k2-3 が指定されても、sortk コマンドは、第 2 欄から第 3 欄までにある語句を比較します。

十六進数で入力

コンマの十六進数は、2Ch です。-t,ではなく-t¥x2C が指定されても、sortk コマンドは、区切り文字としてコンマを使用します。

.csv の拡張子

何も指定されなければ、sortk コマンドは、区切り文字として空白(20h)を利用します。行が並べ替えられるテキストファイルの拡張子が.csv である場合、sortk コマンドは、、自動的に、区切り文字としてコンマを使用します。拡張子の点検を抑制するには、sortk コマンドのパラメータとして、-x-を付加してください。

```

C:¥Q>sortk -t, -k10,20 -o text.tmp Field-20.csv
C:¥Q>sortk -t¥x2C -k10-20 -o yen_x2C.csv Field-20.csv
C:¥Q>sortk -k10-20 -o extension.csv Field-20.csv
C:¥Q>FC text.tmp yen_x2C.csv
ファイル text.tmp と YEN_X2C.CSV を比較しています
FC: 相違点は検出されませんでした
C:¥Q>FC text.tmp extension.csv
ファイル text.tmp と EXTENSION.CSV を比較しています
FC: 相違点は検出されませんでした
C:¥Q>

```

要約および詳細のブロック

利用者は、多数の欄を選択できます。sortk コマンドは、CSV の表を読み込んで、利用者選択の欄で要約を構成して、要約に詳細を結合します。要約のみ比較しながら要約および詳細のブロックを並べ替えます。詳細のみ標準出力に書き出します。

要約を構成

たとえば、21 欄もある CSV で、まず、第 1 欄から第 5 欄までを選択するには、sortk コマンドのパラメータとして、-a5 を付加してください。次に、第 2 欄を選択から除外するには、sortk コマンドのパラメータとして、-h10111 を付加してください。欄を選択することの真偽を示すために、1 または 0 を入力してください。10111 のうち 0 は、第 2 欄を選択しないことを意味します。結果として、4 欄で要約を構成します。ただし、-h10111 が指定されない場合、sortk コマンドは、-h11111 が指定されたとみなします。

-h10111					
-h	1	0	1	1	1
真偽	第 1 欄選択	第 2 欄を除外する	第 3 欄選択	第 4 欄選択	第 5 欄選択

語句の連結

利用者は、多数の欄を選択できます。利用者選択の欄で要約を構成するとき、語句を連結する文字として、ベル文字 (07h) が使用されます。語句連結文字は、sortk の内部で要約に使用されますが、詳細として書き出されません。二桁の十六進数で語句連結文字を設定できます。

制御文字と空白を比較

07h ではなく、たとえば 1Fh を使用するには、sortk コマンドのパラメータとして、-v¥x1F を付加してください。下記の例では、まず、-v¥xHH (H は、十六進数です。[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F]) のパラメータを付加しない sortk コマンドを実行しました。次に -v¥x7F を付加してみました。

```
C:¥Q>sortk -t, -k1,3 cities.csv
Kyoto,Kyoto,Japan
Kyoto City,Kyoto Prefecture,Japan
Nara,Nara,Japan
Nara City,Nara Prefecture,Japan
Osaka,Osaka,Japan
Osaka City,Osaka Prefecture,Japan

C:¥Q>sortk -t, -k1,3 -v¥x7F cities.csv
Kyoto City,Kyoto Prefecture,Japan
Kyoto,Kyoto,Japan
Nara City,Nara Prefecture,Japan
Nara,Nara,Japan
Osaka City,Osaka Prefecture,Japan
Osaka,Osaka,Japan

C:¥Q>
```

語句が短いほど上に来るべきだ

たとえば、小文字の z は、7Ah です。7Fh は、7Ah より大きいいため、Kyoto City の下に Kyoto が来ました。語句が短いほど上に来るように、07h, 1Fh など、20h より小さい数を設定するのが好適です。

引用符を変更したい

引用符が出現したら、sortk コマンドは、もうひとつの引用符が出現するまで、区切り文字を無視しますが、自由な表現として、いわゆる引用符(")が欄内に出現する場合、**sortk コマンドが認識する引用符**を他の文字に変更しなければなりません。

引用符を設定

チルダ(~)が出現した場合もうひとつのチルダが出現するまで区切り文字を無視するように設定するには、sortk コマンドのパラメータとして、-w~ を付加してください。sortk コマンドは、引用符として~を使用します。いわゆる引用符(")は、もはや引用符として認識されません。

```
C:\¥Q>sortk -t, -k2,2 Mark-14.txt
35,~And going a little farther, he fell on the ground and prayed that, if it wer
e possible, the hour might pass him.
36,~And he said, "Abba, Father, all things are possible for you. Remove this cup
 from me. Yet not what I will, but what you will."~
34,~And he said to them, "My soul is very sorrowful, even to death. Remain here
 and watch."~

C:\¥Q>sortk -t, -k2,2 -w~ Mark-14.txt
35,~And going a little farther, he fell on the ground and prayed that, if it wer
e possible, the hour might pass him.
34,~And he said to them, "My soul is very sorrowful, even to death. Remain here
 and watch."~
36,~And he said, "Abba, Father, all things are possible for you. Remove this cup
 from me. Yet not what I will, but what you will."~
```

語句が短いとみなされた

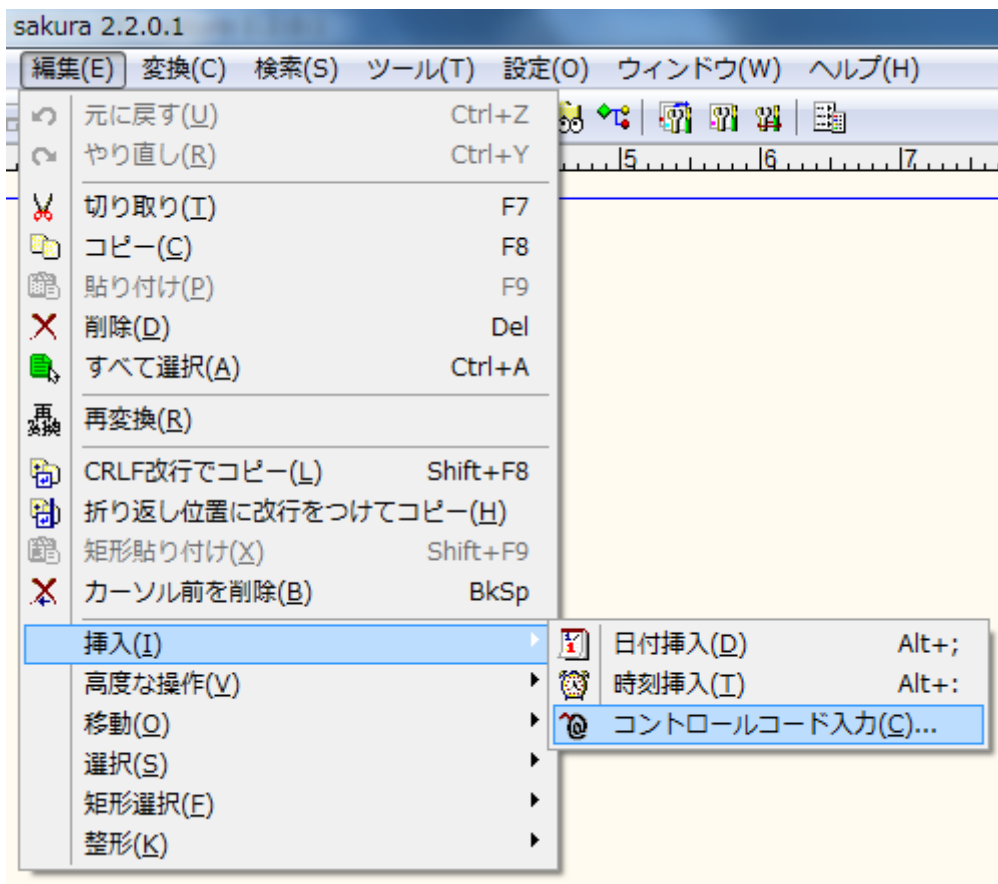
上記の例では、まず、`-wA`(`A` は、ASCII の文字です)のパラメータを付加しない `sortk` コマンドを実行しました。次に`-w~`を付加してみました。`-w~`が付加されていない場合、`sortk` コマンドは、`And he said` の次も区切り文字としてコンマを使用して、語句が短いほど上に来るように行を並べ替えます。このため、マルコ 14 章 36 節がマルコ 14 章 34 節の上に来ました。

いわゆる引用符もコンマも無視するべきだ

自然な表現として、引用符もコンマも `sortk` コマンドに無視されるべきである場合、`-w~` を付加してください。CSV であるにもかかわらずコンマを無視したい範囲を`~`で囲んでください。

制御文字

多機能のテキストエディタは、制御文字を入力する機能があると思われます。たとえば、サクラエディタの場合、編集メニューから挿入を選択して、サブメニューからコントロールコード入力を選択できます。



タブおよび改ページ

タブ、改ページ、エスケープシーケンスなどの制御文字を無視して比較するには、sortk コマンドのパラメータとして、`-i` を付加してください。制御文字の例

十六進数	数値	英語	略号	説明
07h	7	alert	BEL	ベル、警告
08h	8	back space	BS	バックスペース、削除
09h	9	horizontal tab	HT	タブ、水平タブ
0Ah	10	line feed	LF	改行
0Bh	11	vertical tab	VT	垂直タブ
0Ch	12	form feed	FF	改ページ
0Dh	13	carriage return	CR	復帰
1Ah	26	substitute	SUB	MS-DOS テキストファイルの終わり
1Bh	27	escape	ESC	エスケープシーケンスの開始

制御文字を無視するべきだ

下記の例では、mini.txt という CSV にタブ、改ページ、エスケープシーケンスが含まれています。`-i` 無しの `sortk` コマンドは、文字または数字を制御文字と比較します。`-i` 無しの `sortk` コマンドの結果として、40 より上に 86 が来ました。01 より上に 94 が来まし

た。しかし、小さい数が上に来るように、制御文字を無視するべきです。

シフト JIS のテキストファイル

シフト JIS のテキストファイルである場合、`-i` 付きの `sortk` コマンドは、2 バイトの文字、ASCII、半角カタカナのみで要約を構成します。すなわち、1 バイトの文字として、制御文字は、要約に含まれません。

浮動小数点数

浮動小数点数を比較するには、`-g` を付加してください。10 欄まで選択できます。

桁数が同一ではない

桁数が同一ではない整数を比較するには、`-n` を付加してください。20 欄まで選択できます。`-n` が指定された場合、`sortk` コマンドは、もはや語句を比較できません。

```
C:¥Q>sortk -t, -k2,2 -n mini.txt
4956回,01,
5035回,12,
4908回,14,
5043回,15,
5000回,17,
4976回,18,
4944回,19,
4972回,20,
4429回,24,
4960回,25,
4840回,27,
4759回,28,
5049回,32,
4851回,38,
5006回,39,
5014回,40,
4973回,41,
4967回,51,
4902回,59,
4647回,64,
5025回,68,
5030回,70,
5009回,74,
5032回,83,
4881回,86,
4983回,90,
4954回,93,
5037回,94,
C:¥Q>
```

桁数が同一である

下記の例では、mini.txt という CSV にタブ、改ページ、エスケープシーケンスが含まれています。制御文字を無視するために-i を付加した sortk コマンドは、整数でも浮動小数点数でもなく語句を比較します。すなわち、数ではなく数字を比較します。比較対象の数が、すべて 2 桁である場合、-i を付加した例を示します。

```
C:¥Q>sortk -t, -k2,2 -i mini.txt
4956回,01,
5035回,12,
4908回,14,
5043回,15,
5000回,17,
4976回,18,
4944回,19,
4972回,20,
4429回,24,
4960回,25,
4840回,27,
4759回,28,
5049回,32,
4851回,38,
5006回,39,
5014回,40,
4973回,41,
4967回,51,
4902回,59,
4647回,64,
5025回,68,
5030回,70,
5009回,74,
5032回,83,
4881回,86,
4983回,90,
4954回,93,
5037回,94,
C:¥Q>
```

欄の前半にある空白

空白は、制御文字ではありません。比較開始欄の前半にある空白を無視するには、sortk コマンドのパラメータとして、-b を付加してください。下記の例では、添付の atnd コマンドでファイルの先頭から 3 行をコマンドプロンプトに表示しました。空白が無視されたため、並べ替えた結果が異なります。空白を無視するために-b を付加した sortk コマンドは、整数でも浮動小数点数でもなく語句を比較します。

```
C:¥Q>sortk -t, -k1.3 isbn-r.csv | atnd -h3
71,2014,21.00mm,ISBN 978-0-444-63425-2,Elsevier,Dopamine,,By,Marco Diana,
73,2000,38.58mm,ISBN 0-19-853768-9,Oxford,Temporal Logic,,By,Dov M. Gabbay,
73,2010,21.77mm,ISBN 978-0-262-51311-1,MIT,What Is Addiction?,,By,Don Ross,

C:¥Q>sortk -t, -k1.3 -b isbn-r.csv | atnd -h3
100,1993,28.08mm,ISBN 0-8493-4441-7,CRC,Limbic Motor Circuits and Neuropsychiatr
y,,By,Peter W. Kalivas,
100,2010,12.30mm,ISBN 978-0-7618-5041-0,Hamilton,Come Along,We Are Truth-Bound,B
y,Jacinta Respondowska,
100,2010,33.85mm,ISBN 978-1-84882-127-9,Springer,Atlas of Epilepsies,Volume 1,By
,C. P. Panayiotopoulos,

C:¥Q>
```

小数点が発見されない

浮動小数点数を比較するために `-g` が付加された `sortk` コマンドは、小数点を発見しません。小数点が発見されない場合、0.0 になります。

0.0

数字に到達するまで無視される

浮動小数点数を比較するために `-g` が付加された `sortk` コマンドは、比較開始欄の前半にある空白を無視します。負号、小数点、数字に到達するまで、空白だけではなく文字を無視します。整数を比較するために `-n` が付加された `sortk` コマンドは、比較開始欄の前半にある空白を無視します。負号、数字に到達するまで、空白だけではなく文字を無視します。

制限

32 ビット版 Windows において、C++ で `sortk` コマンドを構築しました。ソースコードを添付しました。用途によっては、`sortk` コマンドの利用者は、限界、制限を感じるかもしれません。

要約のサイズ

利用者は、多数の欄を選択できます。`sortk` コマンドは、利用者選択の欄で要約を構成します。`sutai.h` というヘッダーファイルによれば、要約に含める浮動小数点数は、最高 10 個です。1 個の浮動小数点数は、8 バイトです。すなわち、`sortk` コマンドは、80 バイトまで正しく比較します。

比較される語句のサイズ

`sortk` コマンドは、最大 79 バイトまで比較して行の順序が確定することを想定していません。

要約に ID を付加できる

80 バイトまで合致した場合も順序を確定するために、要約に ID を付加するには、sortk コマンドのパラメータとして、-#を付加してください。要約に ID を付加した効果については、後述します。

segmented.h

利用者は、多数の欄を選択できますが、segmented.h というヘッダーファイルによれば、欄数の上限は、3072 です。欄数の上限を変更するには、segmented.h を修正して、sortk を再構築してください。ただし、欄数が多すぎると、sortk コマンドの実行が低速になるおそれがあります。

ヘッダーファイル	名前	数値	説明	問題
sutai.h	BLOCK_SIZE	3072	要約および詳細のブロックのサイズ	行が長すぎるため、詳細を完全には記憶できない。
	DOUBLE_ARRAY_SIZE	10	要約に含める浮動小数点数の個数	要約の上限が 10 欄しかない。10 欄が合致する。行を完全には比較できない。
	INTEGER_ARRAY_SIZE	20	要約に含める整数の個数（浮動小数点数の 2 倍の個数）	要約の上限が 20 欄しかない。20 欄が合致する。行を完全には比較できない。
	END	80	要約に含める 1 バイト文字の個数（浮動小数点数の 8 倍の個数）	要約の上限が 79 バイトしかない。79 バイトが合致する。 行を正しく比較したとはいえない。
	LIMIT_OF_MY_MEMORY	128	要約が合致した行の数	記憶力に限界がある。
segmented.h	FIELD_COUNT	3072	欄数の上限	3072 欄しか選択できない。

大文字と小文字を区別しない

大文字と小文字を区別しないで比較するには、sortk コマンドのパラメータとして、-f を付加してください。ただし、ASCII のみ効果があります。

行数

ファイルの行数は、12800 を超えてはなりません。行数の上限を変更するには、sortk コマンドのパラメータとして、-U 数を付加してください。たとえば、12800 ではなく、131072 を上限とするには、-U131072 を付加してください。下記の例では、-U20 を付加してみました。atnd コマンドで 5156 行もあることがわかりましたが、第 1 行から第 20 行までのみ並べ替えることができました。

```
C:\¥Q>atnd -LF -p20 numbers7.csv
numbers7.csv: 5156 lines
maximum LF-to-LF span of 13 bytes (line-1000)
20,502,2264

C:\¥Q>sortk -t, -k2,2 -U20 numbers7.csv
15,019,6419
4,105,0171
16,179,0524
1,191,1149
3,194,0097
13,229,4673
8,234,7507
9,243,2664
12,247,9722
18,291,3062
11,326,0254
10,340,2436
17,475,8697
20,502,2264
5,592,0265
7,708,3950
14,743,1409
6,792,5482
19,852,1013
2,988,7921

C:\¥Q>
```

改行に到達しないほど小さいバッファ

sortk コマンドは、`-z` 数でバッファのサイズを設定できます。どの行も末尾まで読み込めるように、バッファのサイズを設定してください。極端に小さいサイズが指定された場合、sortk コマンドは、1 回の読み込みで改行に到達しないため、改行文字が含まれていないエラーを表示して中止します。バッファのサイズが指定されていない場合、sortk コマンドは、自動的に最長の行より大きいサイズにバッファを設定します。

EOF の前で改行してください

各行は、改行文字が必要です。**ファイルの終わり(EOF)の直前で改行してください。**制御文字として、`1Ah` をテキストファイルの途中に挿入してはなりません。

diyo コマンドと連携する

添付の diyo コマンドは、MS932~EUC-JP.bin という変換表を編集できます。変換表は、バイナリファイルです。

バージョン番号

短い説明を表示するには、コマンドプロンプトに下記のコマンドおよびパラメータを入力して、Enter を押してください。

diyo -VC

```
C:¥Q>diyo -VC
diyo C++ Version 1.16

diyo -f bars.txt      # Edit according to bars.txt
diyo -kナ -aN        # ナ is substituted with N.
diyo -kナ -aN file   # ナ is substituted with N to produce a binary file
diyo -s① -e1        # ① is substituted with 1 and represented in EUC
diyo -s① -e1 file   # ① is substituted with 1 to produce a binary fi
diyo -s① -e1 -d     # Do not search a table.
diyo -s① -e1 -w     # -w shows which table diyo uses.
```

パラメータ	密接に関係するファイル	説明
-d	MS932~EUC-JP.bin	変換表を探索してはならない。
-f	代用文字のリスト。 MS932~EUC-JP.bin	多数の代用文字にわたって変換表を編集します。
-k ナ	MS932~EUC-JP.bin	ある半角カタカナを廃止します。
-aN	MS932~EUC-JP.bin	ASCII の 1 文字で代用します。
-s①	MS932~EUC-JP.bin	ある機種依存文字を廃止します。
-e1	MS932~EUC-JP.bin	EUC-JP の 1 文字で代用します。
-w	MS932~EUC-JP.bin	どの変換表を使用するか表示します。

代用文字のリスト

事前に利用者が代用文字のリストを用意した場合、diyo コマンドは、代用文字のリストにしたがって、一度に多数の文字を編集できます。

通常文字および代用文字

diyo コマンドは、代用文字リストのファイルから 906 行まで読み込むことができます。代用文字リストのファイルの各行は、通常文字と、代用文字とからなります。利用者は、代用文字もシフト JIS で記入できます。変換表にしたがって、diyo コマンドは、シフト JIS から EUC-JP に利用者記入の文字を変換します。

ファイル名を指定

利用者が代用文字リストのファイルを用意した場合、diyo コマンドのパラメータとして、-f を付加して、空白で区切って、さらにファイル名を付加してください。

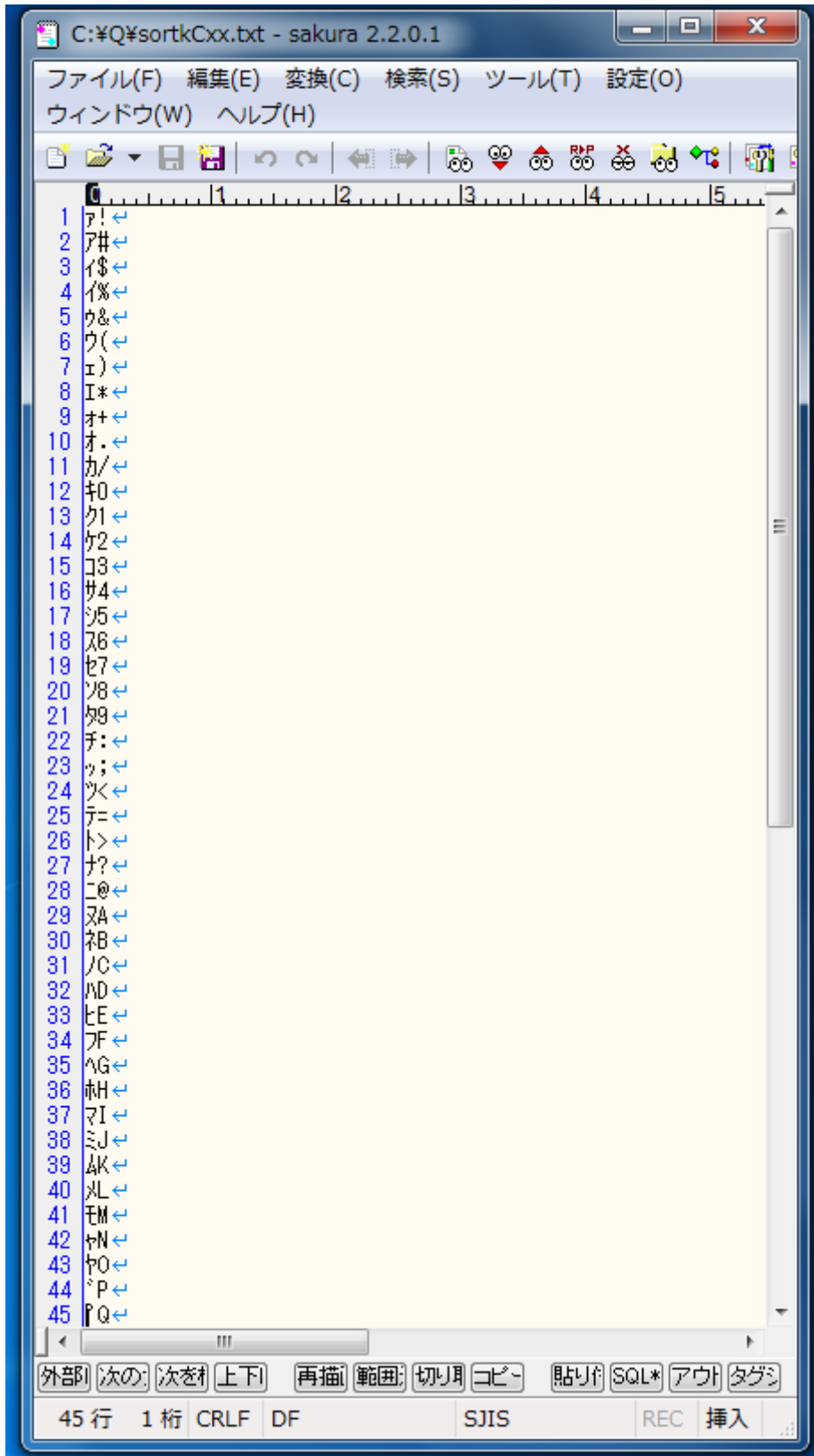
diyo -f bars.txt -w

並べ替えに適した代用文字リスト

上記の bars.txt では、ソフトウェアで並べ替えた結果が昇順になりません。代用文字リストのファイルを作成するとき、並べ替えも考慮したい。

sortk コマンドに適した代用文字リスト

機種(キシユ)、著作権(チョサクケン)などの拗音を直音の付近に並べ、実行可能(ジッコウカノウ)、接頭辞(セツウジ)などの促音を直音の付近に並べ、合図(アイズ)、英語(エイゴ)などの濁音を清音の付近に並べ、添付(テンプ)などの半濁音を清音の付近に並べるのに適した代用文字リストが必要です。C++版の sortk コマンドに適した代用文字リストは、sortkCxx.txt です。Cxx は、C++を意味します。半角カタカナの濁点を廃止して、P で代用します。



0Ah のみで改行できる

EUC-JP のテキストファイルを並べ替えて、0Ah のみで改行するには、sortk コマンドのパラメータとして、-e を付加してください。-8 を付加しても同じ結果を生じます。

第 2 欄を比較

sortk コマンドは、たとえば、第 2 欄にある語句を比較して、CSV のテキストファイルの行を並べ替えることができます。第 2 欄にある語句を比較するには、sortk コマンドのパラメータとして、-k2,2 を付加してください。-k および数字の間に空白を挿入しないでください。半角カタカナの代用として ASCII を使用している例を示します。sortkCxx.txt にしたがって、diyo コマンドが変換表を編集した例(-p については、後述します)

```
C:\>diyo -f sortkCxx.txt -w
C:\>sjisEUCjpCxx~MS932~EUC-JP.bin is primarily accessible to diyo

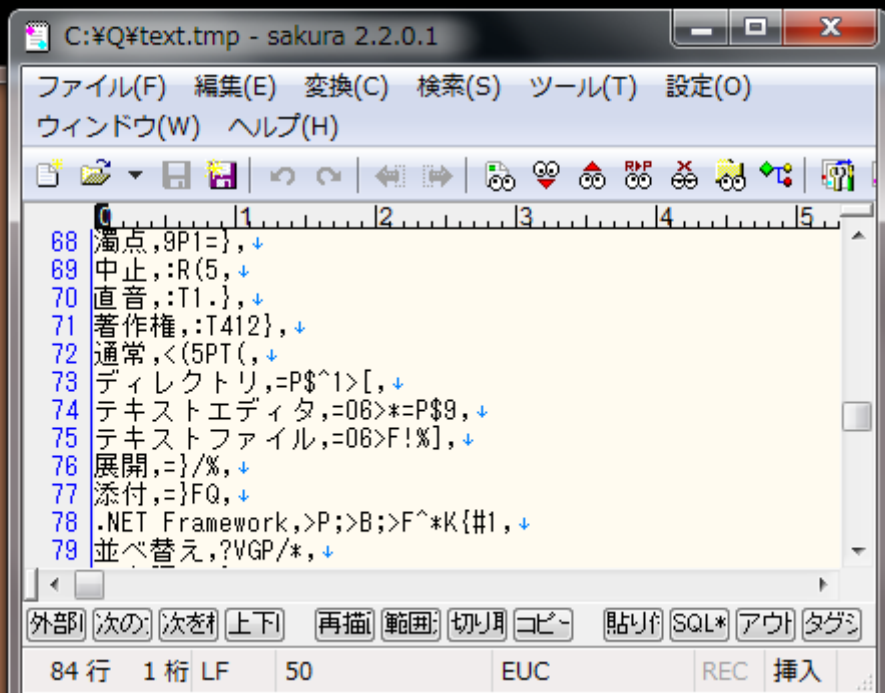
C:\>DIR | FINDSTR "MS932~EUC-JP.bin"
2019/05/15  16:53                31,528 MS932~EUC-JP.bin

C:\>COPY MS932~EUC-JP.bin C:\sjisEUCjpCxx /Y
1 個のファイルをコピーしました。

C:\>sjiseuc Keywords.csv Keywords_euc-jp.csv

C:\>sortk -t, -k2,2 -c- -p -e -o text.tmp Keywords_euc-jp.csv

C:\>
```



濁点を無視

半角カタカナは、清音のカナおよび濁点を組み合わせることで濁音を表現しますが、

清音および濁音は、ほぼ同等の位置になるように並べ替えたい。上記の例のとおり、sortk コマンドが、濁点、半濁点を無視するように、sortk コマンドのパラメータとして、-c-を付加してください。diyo コマンドで、半角カタカナの代用として ASCII を使用している場合、さらに、-p を付加してください。

新しいファイルを命名

新しいファイルを作成するには、sortk コマンドのパラメータとして、-o text.tmp を付加してください。text.tmp は、新しいファイルの名前です。

五十音順

まず、diyo コマンドは、sortkCxx.txt にしたがって変換表を編集しました。次に、sjiseuc コマンドは、シフト JIS から EUC-JP にテキストファイルを変換しました。半角カタカナの読み方は、ASCII に置換されました。さらに、sortk コマンドは、第 2 欄で語句を比較して行を並べ替えました。たしかに五十音順に並べ替えることができました。たとえば、ディレクトリは、テキストエディタの上に来ました。

0Ah のみで改行

0Ah のみで改行したため、サクラエディタの改行は、左向きではなく下向きの矢印で表現されています。

設定ファイル

diyo の設定ファイルは、setting.txt です。利用者がカレントディレクトリから setting.txt を削除しない限り、diyo は、setting.txt にしたがって設定します。setting.txt は、6 行からなります。第 1 行にパス名ではなくファイル名が記入されており、第 4 行の設定内容が空である例

行番号	設定内容	コメント	説明
1	MS932~EUC-JP.bin	table	新しい変換表のファイル名 またはパス名
2	㊿	normal	機種依存文字。通常文字
3	廿	substitute	代用文字
4		bars.txt	代用文字のリスト
5	0	-d	変換表を探索しない
6	0	-w	どの変換表を使用するか表示する

新しい変換表のパス名

diyo が setting.txt を開くことができた場合も、設定を diyo コマンドのパラメータで変更できます。**コマンドがインストールされたディレクトリにある変換表**を新しい変換表のパス名として記入した例

```
C:¥Q>TYPE setting.txt
C:¥s jisEUCjpCxx¥MS932~EUC-JP.bin      table
②①                                         MS932
廿                                           substitute
sortkCxx.txt                             bars.txt
0                                           -d
-W                                         -W

C:¥Q>diyo
C:¥s jisEUCjpCxx¥MS932~EUC-JP.bin is primarily accessible to diyo

C:¥Q>sjiseuc KeyWords.csv KeyWords_euc-jp.csv
C:¥Q>
```

1	コマンド,3I}>P,+
2	シフトJIS,5F>5P6,+
3	テキストファイル,=06>F!%],+
4	下線,/7},+
5	Windows,(\$}>P(6P,+
6	UNIX,S@;16,+

変換表が上書きされる

コマンドがインストールされたディレクトリにある変換表が新しい変換表のパス名として記入されている場合、diyo コマンドは、利用者に警告しないで変換表を上書きします。

利用者が上書きする

setting.txt の第 1 行にパス名ではなくファイル名が記入されている場合、diyo コマンドは、カレントディレクトリに新しい変換表を作成します。diyo がアクセスできるディレクトリに MS932~EUC-JP.bin をコピーしてください(上書きしてください)。

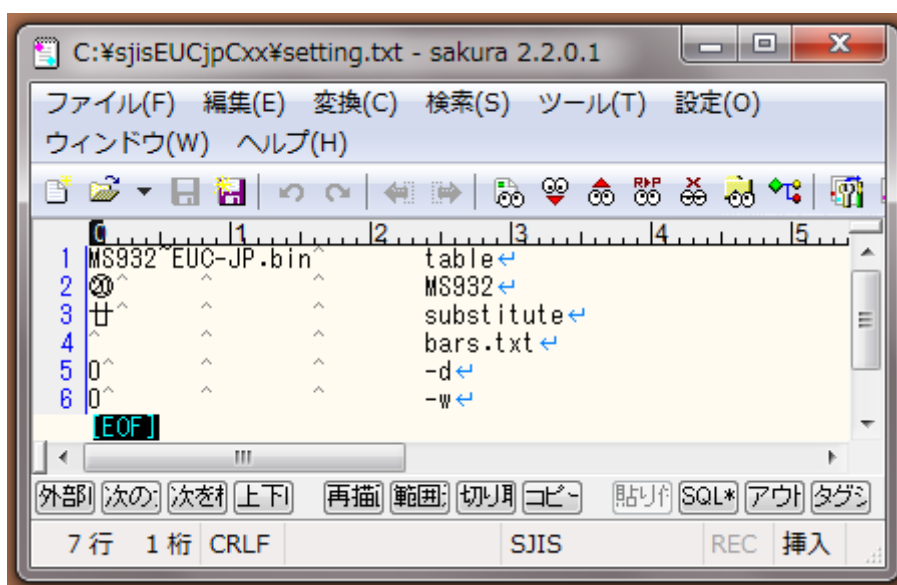
設定ファイルは必須ではない

setting.txt の第 1 行に新しい変換表のパス名またはファイル名を記入できますが、設定ファイルを開くことができない場合、diyo コマンドは、利用者に警告しないで、カレントディレクトリに新しい変換表を作成します。diyo がアクセスできるディレクトリに MS932~EUC-JP.bin をコピーしてください。

内容、タブ、コメントの順に記入

設定内容を左に詰めて記入してください。設定内容のみで改行できますが、シングル

クォーツ、コンマ、縦線、タブの次にコメントを記入できます。Perl, PHP, Python, Ruby などのスクリプト言語とは異なり、diyo は、シャープ(#)の次にコメントを記入できません。空白(20h)の次にコメントを記入できません。空白ではなくタブを利用してください。Windows のメモ帳では、タブが機能しますが空白に見えます。サクラエディタで、タブが脱字記号(^)で表示されている例



0 ではないならば有効

第 5 行、第 6 行に 0 を記入すると無効になります。どの変換表を使用するか表示したくないなら第 6 行に 0 を記入してください。

第 4 行の設定内容が空である

代用文字リストのファイル名が空である場合、多数の代用文字にわたって変換表を編集する機能は、無効になるため、第 2 行および第 3 行にしたがって、1 文字を編集します。

第 4 行に sortkCxx.txt を記入

代用文字リストのファイル名が空ではない場合、第 4 行にしたがって、多数の代用文字にわたって変換表を編集します。

五十音順

下記の例では、第 4 行に sortkCxx.txt を記入しました。まず、diyo コマンドは、sortkCxx.txt にしたがって変換表を編集しました。次に、sjiseuc コマンドは、シフト JIS から EUC-JP にテキストファイルを変換しました。半角カタカナの読み方は、ASCII に置換されました。さらに、sortk コマンドは、第 2 欄で語句を比較して行を並べ替えました。たしかに五十音順に並べ替えることができました。たとえば、ディレクトリは、テキストエ

データの上に来ました。添付のnarfiコマンドは、[第 2 欄を削除できます](#)。

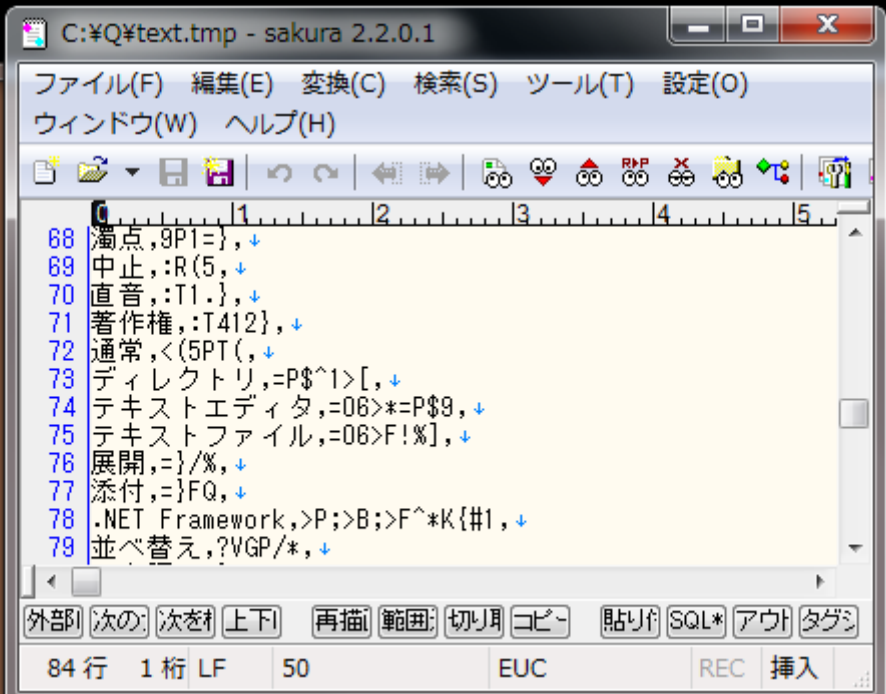
```
C:¥Q>TYPE setting.txt
C:¥s jisEUCjpC¥x¥MS932~EUC-JP.bin          table
②①                                             MS932
卍                                             substitute
sortkC¥x.txt                                  bars.txt
0                                               -d
-w                                              -w

C:¥Q>diyo
C:¥s jisEUCjpC¥x¥MS932~EUC-JP.bin is primarily accessible to diyo.

C:¥Q>sjiseuc KeyWords.csv KeyWords_euc-jp.csv

C:¥Q>sortk -t, -k2,2 -c- -p -e -o text.tmp KeyWords_euc-jp.csv

C:¥Q>
```



The screenshot shows a sakura 2.2.0.1 text editor window with the following content:

```
68 濁点,9P1=},↓
69 中止,:R(5,↓
70 直音,:T1.},↓
71 著作権,:T412},↓
72 通常,<(5PT(,↓
73 ディレクトリ,=P$^1>[,↓
74 テキストエディタ,=06>*=P$9,↓
75 テキストファイル,=06>F!%],↓
76 展開,=}/%,↓
77 添付,=}FQ,↓
78 .NET Framework,>P;>B;>F^*K{#1,↓
79 並べ替え,?VGP/*,↓
```

The editor's status bar at the bottom indicates: 84行 1桁 LF 50 EUC REC 挿入

ID として行番号を使用する

2 行以上が同じ要約を生じても順序が確定するように、要約に ID を付与できます。sortk コマンドは、ID として行番号を使用します。要約も ID も標準出力に書き出されません。

どの行も唯一だ

利用者が、比較対象として欄をどう選択しても、行番号が異なるため、比較対象は、どれも唯一です。下記の例では、まず、-#を付加しない sortk コマンドを実行しました。次に、要約に ID を付加するために-#を付加して sortk コマンドを実行しました。第 10

欄から第 20 欄まで比較して順序が確定しない場合も、行番号で順序が確定します。

```
C:\%Q>sortk -t, -k10,20 Field-22.csv | atnd -h23
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,07,12,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,08,8,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,10,13,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,1,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,22,15,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,26,17,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,27,10,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,29,18,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,44,11,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,49,7,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,51,2,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,59,23,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,59,24,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,59,25,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,59,26,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,59,27,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,59,28,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,59,29,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,59,30,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,59,31,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,59,32,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,59,22,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,59,20,
C:\%Q>sortk -t, -k10,20 -# Field-22.csv | atnd -h23
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,07,12,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,08,8,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,10,13,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,1,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,22,15,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,26,17,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,27,10,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,29,18,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,44,11,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,49,7,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,51,2,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,59,19,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,59,20,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,59,21,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,59,22,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,59,23,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,59,24,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,59,25,
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,59,26,
```