

# 文字コード掲示板

[\(ホームへ\) 戻る\(C\)](#) [記事一覧\(L\)](#) [新規投稿\(N\)](#) [全文検索\(S\)](#) [管理\(U\)](#)

## 記事表示

**093 Java の MS932, Cp943C, SJIS の違い** <森山 将之> 2003/08/26(火) 23:44

No.093 投稿日時: 2003/08/26(火) 23:44 <親記事>

投稿者: 森山 将之

### Java の MS932, Cp943C, SJIS の違い

Java※ の MS932, Cp943C, SJIS の変換で異なる点、および注意を要する点をまとめてみました。

※調査したバージョン: Java(TM) 2 Runtime Environment, Standard Edition (build 1.4.1\_02-b06)

#### ■概要

MS932 と Cp943C の両者は、Windows-31J の文字セットを扱えます。

主な違いは、Unicode への変換で一部異なるコードポイントに変換される事と、NEC特殊文字とIBM拡張文字の両方で定義されている文字を、Unicode から MS932/Cp943C に変換する際に、どちらのコードポイントかという点が異なります。

MS932/Cp943C は Unicode との対応付けが一部の JIS X 0208 の文字に関して、SJIS と異なり、EUC\_JP や ISO2022JP へ変換出来ない文字があるので注意が必要となります。

#### ■MS932 と Cp943C で異なる Unicode コードポイント

MS932 と Cp943C では、Unicode に変換した時に 表1 に示した違いがあります。

表1 MS932 と Cp943C で異なる Unicode のコードポイント

文字	シフト JIS	Unicode	
		MS932	Cp943C
—	815C	U+2015	U+2014
～	8160	U+FF5E	U+301C
	8161	U+2225	U+2016
-	817C	U+FF0D	U+2212
¡	FA55	U+FFE4	U+00A6

次の 表2 は、Unicode のコードポイントは同一だが、MS932/Cp943C への変換で異なるコードポイントへ変換される文字の一覧です。

表2 Unicode → MS932/Cp943C で変換先が異なるコードポイント

文字	MS932 or Cp943C (シフト JIS)		→Unicode	→MS932	
	NEC特殊文字	IBM拡張漢字		NEC特殊文字	→Cp943C IBM拡張漢字
I	8754	FA4A	U+2160	8754	FA4A
II	8755	FA4B	U+2161	8755	FA4B
III	8756	FA4C	U+2162	8756	FA4C
IV	8757	FA4D	U+2163	8757	FA4D
V	8758	FA4E	U+2164	8758	FA4E
VI	8759	FA4F	U+2165	8759	FA4F
VII	875A	FA50	U+2166	875A	FA50
VIII	875B	FA51	U+2167	875B	FA51
IX	875C	FA52	U+2168	875C	FA52
X	875D	FA53	U+2169	875D	FA53
No.	8782	FA59	U+2116	8782	FA59
Tel	8784	FA5A	U+2121	8784	FA5A
綈	878A	FA58	U+3231	878A	FA58

※MS932 では、NEC特殊文字に変換され、Cp943C では、IBM拡張文字に変換されます。

#### ■JIS X 0208-1983 で追加された文字に変換されるコードポイント

NEC特殊文字、NEC選定IBM拡張文字、IBM拡張文字で定義され、後に JIS X 0208-1983 で 2 区に追加になった文字は、JIS X 0208 の 2 区のコードポイントに変換されます。

表3 は、その一覧です。

表3 JIS文字に変換されるコードポイント

文字	MS932 or Cp943C (シフト JIS)			→ Unicode	→ MS932 or Cp943C JIS X 0208 2 区
	NEC特殊文字	NEC選定 IBM拡張文字	IBM拡張文字		
U	879C			U+222A	81 BE
∩	879B			U+2229	81 BF
∠	8797			U+2220	81 DA
⊥	8796			U+22A5	81 DB
≡	8791			U+2261	81 DF
≋	8790			U+2252	81 E0
√	8795			U+221A	81 E3
∴	879A		FA5B	U+2235	81 E6
§	8792			U+222B	81 E7
ㄣ		EEF9	FA54	U+FFE2	81 CA

#### ■MS932/Cp943C と SJIS の相違点

MS932 もしくは Cp943C コンバーターは Unicode との対応付けで、表4 の赤字で示したコードポイントが、SJIS コンバーターで変換した場合と異なります。

EUC\_JP や ISO02202JP コンバーターは、SJIS コンバーターと同じ Unicode のコードポイントを用いていますので、932/Cp943C との相互変換の際には注意が必要となります。

す。

文字	シフト JIS	Unicode		
		SJIS	MS932	Cp943C
ー	815C	U+2014	U+2015	U+2014
～	8160	U+301C	U+FF5E	U+301C
	8161	U+2016	U+2225	U+2016
ー	817C	U+2212	U+FF0D	U+2212
⌘	8191	U+00A2	U+FFE0	U+FFE0
⌘	8192	U+00A3	U+FFE1	U+FFE1
⌘	81CA	U+00AC	U+FFE2	U+FFE2

### ■重複符号化されている文字

MS932/Cp943C と、Unicode への変換で、重複符号化されている文字は、多対1 の変換となっています。

Unicode から MS932/Cp943C への変換は、表2 の文字以外は、MS932 と Cp943C は同一の変換となっています。

MS932 の変換は、次のページをご覧ください。

Windows-31J の重複符号化文字と Unicode (当サイト・コンテンツ)

<http://www2d.biglobe.ne.jp/~msyk/charcode/cp932/uni2sjis.html>

### ■ユーザー定義文字

MS932/Cp943C では、95～114区 (F040～F9FC) がユーザー定義文字の領域となっています。

両方のコンバーターとも Unicode の私用領域の U+E000～U+E757 と対応付けされています。

### ■参照

WebSphere V3.5 発表ワークショップ資料

> 第3回「プログラミング・モデル(JSP、Servlet)」> 文字コード

[http://www-6.ibm.com/jp/software/websphere/developer/wsv35wslib/pdf/was35\\_psj5\\_1.pdf](http://www-6.ibm.com/jp/software/websphere/developer/wsv35wslib/pdf/was35_psj5_1.pdf)

Windows-31J情報 (当サイト・コンテンツ)

<http://www2d.biglobe.ne.jp/~msyk/charcode/cp932/index.html>

文字コード掲示板 過去ログ

<http://www2d.biglobe.ne.jp/~msyk/cgi-bin/charcode/bbs.cgi?c=gr&n=55>

※間違いや不明な点がありましたらご指摘ください。よろしく願いいたします。

