

Shift JIS-2004 は、JIS X 0213 の符号化方式のひとつです。JIS X 0213 の符号を計算式によって変形したものと JIS X 0201 とを同時に用いる方式で、1 バイトと 2 バイトの混在する 文字コード です。

JIS X 0213 の附属書 1 で定義されています。JIS X 0208 ベースの Shift JIS の上位互換となっています。このため、Shift JIS-2004 に対応すれば Shift JIS にも対応したことになります。

符号の構造

Shift JIS と同じく、JIS X 0201 の 8 ビットコードの隙間に JIS 漢字 の 2 バイトコードを変形のうえ詰め込んだものです。

1 バイトコードの範囲は、0x00-0x7F (JIS X 0201 ラテン文字集合)、0xA0-0xDF (JIS X 0201 片仮名集合) です。

2 バイトコードの第 1 バイトとなる範囲は、0x81-0x9F, 0xE0-0xFC です。

第 1 バイトが 0xEF までの範囲は、漢字集合 1 面 (非漢字ならびに第 1 ~ 第 3 水準漢字) に相当し、従来の Shift JIS と同じ計算方法で区点番号から符号化表現 (第 1・第 2 バイト) を求めることができます。

第 1 バイトが 0xF0 から 0xFC までの範囲は、漢字集合 2 面 (第 4 水準漢字) に相当します。区点番号から符号化表現を求めるには、異なる計算式を用いる必要があります。計算方法は下記リンク先を参照してください。

- ・ JIS X 0213 の代表的な符号化方式 #Shift JIS-2004

名称について

Shift JIS-2004 という名称になったのは JIS X 0213:2004 においてです。規格初版の JIS X 0213 :2000 では、Shift JISX0213 という名称でした。

Shift JISX0213 と Shift JIS-2004 とでは、JIS X 0213 の 2000 年版と 2004 年版の違い、すなわち 10 文字追加による差異があります。しかし全体に対して見ればごくわずかな違いであるため、両者を区別しないことがしばしばあります。

実装

Mac OS X や、プログラミング言語の Python, Java, PHP, コード変換プログラムの iconv, nkf, テキストエディタの GNU Emacs 等はこの符号化方式に対応しています。

- ・ iconv では、"SHIFT_JISX0213" あるいは "SHIFT_JIS-2004" の名前でこの符号化方式に対応しています
- ・ Mac OS X に付属のテキストエディットや Safari では、"Shift JIS X0213" という名称で対応しています

- ・ Java では "x-SJIS_0213" という名前で指定します (Java 7 以降で有効。 Java 6 でも途中の Update から対応)

例えば、 iconv コマンドで Shift_JIS-2004 から UTF-8 に変換するには下記のようにします。

```
iconv -f SHIFT_JISX0213 -t UTF-8 < sjis.txt > utf8.txt
```

シフト JIS から UTF-8 へコード変換するには常にこの方法で行うのがおすすめです。Windows のベンダ定義外字の丸付き数字やローマ数字も救済でき、かつ、波ダッシュを正しい符号位置 (U+301C) に変換します。 波ダッシュだけでなく、双柱や負符号なども文字化けせず正しく変換します。

UTF-8 から SJIS に変換するには、上の -f と -t を入れ替えれば OK です。

```
iconv -f UTF-8 -t SHIFT_JISX0213 < utf8.txt > sjis.txt
```

参考

- ・ JIS X 0213 の代表的な符号化方式
- ・ JIS X 0213 のコード対応表

関連項目

- ・ Shift_JIS - 本符号化方式はこれの拡張版
- ・ JIS X 0213 - 元になっている規格
- ・ EUC-JIS-2004 - 本符号化方式と同様の JIS X 0213 の符号化方式の一種
- ・ ISO-2022-JP-2004 - 本符号化方式と同様の JIS X 0213 の符号化方式の一種

コメント

- ・ ウィンドウズのシフトジスとは非互換。はしご高や立ざきが文字化けする。 - よよよ (2019年11月19日02時36分58秒)
- ・ サファリのメニューにシフトジスエックス 0213 あるけど日本語は表示できません - よよよ (2019年11月19日02時46分26秒)
- ・ Windows のベンダ定義外字は Windows 側で修正すべき問題です。 CP932 では第3第4水準漢字もアクセント付きアルファベットもアイヌ語も使えません。 - jisx0213 (2021年12月12日16時53分49秒)
- ・ ウィンドウズのシフトジス (=CP932) と互換性を保ちたい場合は CP932X も可。はしご高も立ざきもそのまま使えて、第3第4水準漢字もアクセント付きアルファベットもアイヌ語も使える。 - CP932X (2023年04月07日03時04分48秒)