

EUC-JIS-2004 は、JIS X 0213 の符号化方式のひとつであり、JIS X 0213 と ASCII を同時に運用するものです。JIS X 0213 の附属書 3 で定義されています。EUC-JP から JIS X 0212 補助漢字を除いたものに対して上位互換となっています。EUC-JP の実装には補助漢字をサポートしていないことがよくあるので、多くの場合は単に EUC-JP の上位互換とみなせます。

符号の構造

最上位ビットが 0 になっているバイト (つまり 0x7F 以下のバイト) は、常に ASCII の文字を表します。

0xA1 から 0xFE までの範囲は 2 バイトで JIS X 0213 の漢字集合 1 面の文字を表します。ただし、直前に 0x8F を付けると、2 バイトで漢字集合 2 面の文字を表します。

漢字集合 1 面・2 面どちらの文字も、JIS X 0213 で定義されているコード (ビット組合せ) の各バイトに最上位ビット 1 を付けて表します。Shift JIS-2004 のように計算で変形したりはしません。

0x8E の直後 1 バイトは、最上位ビットを 1 にした状態で JIS X 0201 片仮名集合の文字を表します。

実装

プログラミング言語の Python, Ruby, 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。

Emacs 上の日本語入力プログラム SKK では JIS 第 3 第 4 水準漢字辞書 が整備されていますが、その辞書ファイルには EUC-JIS-2004 が使われています。

また、全国 1 万以上の医療施設で使われている 日医標準レセプトソフト では内部的に EUC-JIS-2004 が使われているとのことです。

コード変換例

例えば、Mac のコマンドラインや Windows の Cygwin で iconv を使って UTF-8 との間でコード変換するには下記のようにします。文字コード名「EUC-JISX0213」によって EUC-JIS-2004 を指定できます。

```
iconv -f UTF-8 -t EUC-JISX0213 < utf8.txt > euc.txt
```

これによって UTF-8 のファイル utf8.txt を EUC にコード変換してファイル euc.txt に書き出します。変換方向を逆にするには -f と -t の指定を入れ替えます。EUC と UTF-8 の間で変換するには、常にこのやり方を用いるのが適当です。これは Windows の実装にある 波ダッシュ問題 を引き起こさず、また 丸付き数字 やローマ数字を救済できます。EUC のコード変換には常にこれを用いるのがおすすめです。

名称について

EUC-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、EUC-JISX0213 という名称でした。

iconv では、EUC-JISX0213 の名前でこの符号化方式に対応しています。Ruby 2.0 では "EUC-JP-2004" という名前で登録されています。Ruby 2.1.3 では "EUC-JIS-2004" に修正されたそうです (参考: Ruby 2.1 で EUC-JIS-2004 の名前が修正)。

他の符号化方式との関係

この符号化方式は、規格本体で定義されている「国際基準版・漢字用 8 ビット符号」に JIS X 0201 片仮名を追加したものです。この片仮名集合は EUC-JIS-2004 では原則として使用しないことになっているので、その原則を適用すれば実質的に「国際基準版・漢字用 8 ビット符号」と同じになります。

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 - Unicode との機械可読形式の変換表があります

関連項目

- ・ EUC-JP - JIS X 0208 の対応する符号化方式
- ・ JIS X 0213 - 元になっている規格
- ・ Shift 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
- ・ 国際基準版・漢字用 8 ビット符号