

Re: ISO-2022-JP-2004 vs Unicode mapping table - えむけい (2006 年 06 月 06 日 06 時 01 分 35 秒)

ISO-2022-JP-2004 の対応表ですが、ISO/IEC 646 IRV と重複する文字も漢字 7 ビット符号の対応表と同様に割り当ててるのではないですか？

重複符号化になりますが、ISO-2022-JP-2004 に関しては (EUC-JIS-2004 や Shift_JIS-2004 と異なり) 使用禁止も「慣用的な利用との互換」も JIS X 0213 の附属書に規定されていません。

つまり FULLWIDTH に割り当ててる根拠がありません (参考として「Fullwidth:」を付けるのはかまわないと思いますが)。

- ・ おや。どうもそのように読めますね。JIS X 0208 の附属書 2 (RFC1468 符号化表現) も同様のよう見えます。これはどう解釈すれば良いのでしょうか。規格の誤り? - やの (2006 年 06 月 07 日 23 時 42 分 45 秒)
- ・ 私は誤りではなく、ISO-2022-JP 的に文字集合を切り替えて使う場合に重複符号化が発生する可能性があるのは仕様なのだとして解釈しています。実際、たとえば 1b 24 42 と 1b 24 28 51 ではほとんどの文字が重複していますが、それを理由にどちらかの使用を禁止されたりはしていません (1b 24 42 の一部の文字は使用禁止ですが、重複符号化を理由としているわけではありません)。SuikaWiki の [図形文字の一意的符号化](#) が参考になるとと思います。 - えむけい (2006 年 06 月 08 日 12 時 47 分 09 秒)
- ・ うーん、すると Unicode に変換したときに「全角・半角」の区別が保存されなくなりますよね。それに、ISO-2022-JP{-3}-2004 で「全角英数」を "FULLWIDTH" の方に変換している実装は全部間違いということになってしまいますけど、それで良いのでしょうか? - やの (2006 年 06 月 08 日 22 時 44 分 50 秒)
- ・ そもそも JIS X 0213 の英数字は「全角」ではないので、区別が保存されないのはそういうものでは? というか。EUC-JIS-2004 と Shift_JIS-2004 ですら同じ JIS X 0213 の文字が「半角」になったり「全角」になったりしてますし。規格を字義どおりに解釈するなら、ISO-2022-JP{-3}-2004 で FULLWIDTH に割り当てて実装は「全部間違い」としかいいようがないです。[XML MOJII00655](#) は私のお気に入りの言葉です。川俣氏が Windows の変換表の誤りに関して [ダブルスタンダード](#) なのは残念というほかありませんが。 - えむけい (2006 年 06 月 09 日 09 時 03 分 54 秒)
- ・ JIS X 0213:2000 の本体の 9. 符号拡張法、9.2 には、「JIS X 0201 又は ISO/IEC646 と漢字集合とを同時に指示する場合、これまでの慣用的な利用との互換を目的としてだけ、附属書 5 表 2 に規定される文字を JIS X 0201 又は ISO/IEC 646 とで規定される文字とは異なった図形文字として用いてもよい。」とあります。しかし附属書 2 から本体 9. を参照できるのかどうかは、よく分かりません。 - さだひろ (2006 年 06 月 10 日 10 時 09 分 43 秒)
- ・ もう少し考えてみましたが、本体 9.2 も附属書 2 も漢字集合として本体 6.5.1 で規定するものを用いる点では共通しています。すると、附属書 2 が、本体 9.2 における「JIS X 0202 の符号拡張法を主とする環境で」に該当するのかどうかなのでしょうが、どうなのでしょう... - さだひろ (2006 年 06 月 10 日 10 時 28 分 00 秒)
- ・ 上で「全角・半角」という言葉を使ったのは便宜のためです。本質的には、同じ文字に異なる符号化表現を重複して割り当てている場合に、コード変換で重複が失われないようにするよう Unicode では配慮されており、それが fullwidth/halfwidth form の意義である筈です(「互換漢字」の存在もその目的)。とすれば、ISO-2022-JP で fullwidth form を使わない理由は存在しないはずだと言うのが私の意見です。規格は字義通り解釈しなければならないと言うのは当然のことですが、規格の誤りを発見し改められるようにすることも大事なことだと思います。 - やの (2006 年 06 月 14 日 23 時 21 分 36 秒)
- ・ えっと、JIS X 0213:2000 において、EUC-JISX0213 が附属書 5 を参照しているのに、ISO-2022-JP-3 が附属書 5 を参照していなかった理由ですけど、これ、ISO/IEC 2022 を厳密に解釈したら、こうなったというだけのことです。EUC-JISX0213 の場合は、G0 と G1 に「同じ名前」の文字 (たとえば LATIN CAPITAL LETTER A) が収録されてるので、そのままだと G1 の 1-3-33 とかが使えなくなっちゃいますよね。で、それが困る人は代替名称にするわけです。でも、ISO-2022-JP-3 は G0 を変えながら使うので、「同じ名前」の文字が同時に G0 には現れない。でするので、ISO/IEC 646 IRV の 4/1 も JIS X 0213 の 1-3-33 もどちらも使えるので、ISO/IEC 2022 上はあえて代替名称を使う理由がないわけです。じゃあ、ISO-2022-JP-3 では代替名称を使っちゃいけないか、っていうと、それはもちろん「慣用的な利用との互換」のためには使ってもかまわないわけで、それがわざわざ附属書 5 を独立させている理由でもあるわけです。 - 安岡孝一 (2006 年 06 月 25 日 22 時 07 分 36 秒)

- ・ あ、そうか、やのさんは、JIS X 0213 での「名前」を、JIS X 0221-1 の「名前」と同一視してらっしゃるんですね。それ、厳密には違いますよ。JIS X 0213 での「名前」は、JIS X 0202 の 5.3 でいう「名前」でもあるんです。その意味で、JIS X 0213:2000 附属書 5 には、あえて UCS を示さなかったんです。ただ、1-1-17 の代替名称をドタンバで「FULLWIDTH OVERLINE」にできなかった（理由は JIS X 0221-1:2001 附属書 P 参照）のは、いまだもってクヤシかったりしますけど。 - 安岡孝一 (2006 年 06 月 26 日 00 時 10 分 27 秒)
- ・ なるほど、附属書 5 は独立しているから附属書 2 から参照しても構わないと。考えてみれば附属書 4 や 6 も参照するわけだから、附属書 5 だけ参照してはいけない理由はないですよ。言われてみれば単純なことでした。 - やの (2006 年 06 月 29 日 22 時 17 分 41 秒)