

概要

JIS X 0208 は、日本の文字コード規格で、JIS X 0213 の元になったものです。規格名称を「7 ビット及び 8 ビットの 2 バイト情報交換用符号化漢字集合」といいます。

ISO/IEC 2022 に整合的な 94 × 94 文字の 2 バイト符号化文字集合として設計されています。漢字をはじめ、平仮名、片仮名、ラテン文字、ギリシャ文字、キリル文字、数字、記号類を含む 6,879 文字のコードを定めています。

1978 年の制定以降、日本語に対応する情報機器において極めて広く利用されてきました。ただし、漢字においても漢字以外においても文字が足りない、現代日本のテキストを符号化するのに不十分であるという問題があり、拡張規格の JIS X 0213 の制定につながりました。

制定・改訂履歴

- ・ 1978 年：初版制定。(当初の規格番号は C6226)
- ・ 1983 年：改正。記号類の追加のほか、一部の漢字の符号位置の入れ替え、字体の簡略化。
- ・ 1990 年：改正。漢字を 2 文字追加。
- ・ 1997 年：改正。コード表には変更を加えず、符号位置に対応する文字の明確化や漢字の典拠調査を行う大規模な改正。包摂規準の明確化、シフト JIS の規定、文字名の規定 (Unicode との変換表に相当) 等。
- ・ 2012 年：改正。2010 年の常用漢字改正による参考情報の更新。技術的な変更はなし。

符号化方式

運用方法(いわゆる文字符号化方式)として、JIS X 0208 を単体で用いる方式(すなわち 2 バイト固定長)のほか、JIS X 0201 ならびに ISO/IEC 646 国際基準版 (ASCII と同等) と同時に用いる方式(すなわち 1 バイトと 2 バイトが混在)を定義しています。また、附属書において シフト JIS と ISO-2022-JP (RFC1468) を取り込んでいます。

よく利用される符号化方式として下記の 3 つがあります。

- ・ シフト JIS (Shift JIS): JIS X 0201 の隙間に JIS X 0208 を変形のうえ詰め込んだコード。JIS X 0208 の附属書 1。
- ・ EUC-JP: ISO/IEC 2022 の 8 ビット符号表の GL 領域に ASCII、GR 領域に JIS X 0208 を呼び出した状態のコード。オプションで JIS X 0201 片仮名、JIS X 0212 を用いることも。ASCII と JIS X 0208 だけの場合は、JIS X 0208 で規定される「国際基準版・漢字用 8 ビット符号」に相当。
- ・ ISO-2022-JP: ASCII と JIS X 0208 を エスケープシーケンス で切り替える 7 ビットのコード。

規格本体で規定される 6 種類の符号化方式は下記の通りです。これらも JIS X 0213 に受け継がれています。

- ・ 漢字用 7 ビット符号
- ・ 漢字用 8 ビット符号
- ・ 国際基準版・漢字用 7 ビット符号 - ISO/IEC 646 国際基準版と漢字集合を SI/SO で切り替え
- ・ 国際基準版・漢字用 8 ビット符号 - ISO/IEC 646 国際基準版を GL に、漢字集合を GR に (EUC-JP とほぼ同じ)

- ・ ラテン文字・漢字用 7 ビット符号 - JIS X 0201 ラテン文字集合と漢字集合を SI/SO で切り替え
- ・ ラテン文字・漢字用 8 ビット符号 - JIS X 0201 ラテン文字集合を GL に、漢字集合を GR に

JIS X 0213 との関係

上記の符号化方式それぞれに対応した拡張版の符号化方式が JIS X 0213 でも用意されています。附属書には Shift JIS-2004, EUC-JIS-2004, ISO-2022-JP-2004 が定義されています。

JIS X 0213 は JIS X 0208:1997 がベースになっています。

JIS X 0213 の漢字集合 1 面は、JIS X 0208 の上位互換として設計されています。

Unicode (ISO/IEC 10646 UCS) との関係

JIS X 0208 は Unicode の CJK 統合漢字 を最初に作る際に参照された原規格のひとつです。

JIS X 0208 の各文字に対応する Unicode の文字は、JIS X 0208 の規格の中に、ISO/IEC 10646 UCS (Unicode と同じ) の文字名を参照する形で規定されています。これに従って Unicode との間でコード変換ができます。この方法は JIS X 0201 や ISO/IEC 646、ISO/IEC 8859 等と同じです。

機械可読な変換表は、JIS X 0213 と Unicode のコード変換表が使えます。JIS X 0213 は JIS X 0208 の上位互換なので、ここにある対応表の中から JIS X 0213 で追加された文字 ([2000] 及び [2004] とマークされた行) を読み飛ばすと、JIS X 0208 の変換表として使えます。

なお、いわゆる「CP932」などのベンダ実装の変換表は、Unicode と JIS の対応関係に誤りが幾つかあるので不適切です。例えば、JIS の波ダッシュ (文字名 WAVE DASH, SJIS 0x8160) は Unicode の U+301C (文字名 WAVE DASH) に変換するのが正しいですが、「CP932」と呼ばれるものでは、誤って U+FF5E (FULLWIDTH TILDE, EUC-JP 等のチルダを重複符号化した文字コードとの間の往復変換のための救済用の符号位置) に対応づける実装が存在します。このような実装は文字化けの元になっており、波ダッシュ問題として知られています。

JIS X 0208 についての誤解

JIS X 0208 は文字の集合だけを定めていて実際にコンピュータ上で表現される符号化形式は定めていないという誤解がたまにあります。また、JIS X 0208 が定めるのは各文字の整数としての区点番号だけだという誤解もあります。実際には区点番号は符号表の座標を指すための番号であり、各区点には 2 バイトのビット組合せが一意に対応します。

正しくは、JIS X 0208 は各文字に対して 2 バイトのビット組合せを定義しています。各バイトは 7 ビットであり、第 8 ビットは、8 ビット符号表の GL 領域 (0x20-0x7F) に呼び出したときには 0、GR 領域 (0xA0-0xFF) のときは 1 となります。また、シフト JIS は JIS X 0208 の符号化表現を計算によって変形したうえで JIS X 0201 と組み合わせた運用方式 (符号化方式) です。

参照情報

JIS X 0208 のコード表は、ISO-IR のサイトで PDF ファイルを参照できます (番号 168)。

規格票は JSA Web Store で購入できます。

参考

- ・ プログラマのための文字コード技術入門 第 3 章

関連項目

- ・ JIS X 0213 - この規格の拡張版。
- ・ Shift JIS - JIS X 0201 とこの規格を同時に用いる符号化方式の一つ。
- ・ EUC-JP - ASCII とこの規格を同時に用いる符号化方式の一つ。
- ・ ISO-2022-JP - ASCII/JIS X 0201 とこの規格を同時に用いる符号化方式の一つ。