

EUC (Extended Unix Code) は、ASCII と 2 バイト文字コードを一緒に運用する方式のひとつです。

8 ビット符号表の GL 領域 (0x20-0x7F) には ASCII を呼び出し、GR 領域 (0xA0-0xFF) には 2 バイトコードを呼び出した状態で使います。2 バイトコードは、例えば JIS X 0208 や中国の GB 2312、韓国の KS X 1001 などです。

日本の EUC-JIS-2004 や EUC-JP、中国の EUC-CN、韓国の EUC-KR、台湾の EUC-TW といった種類があります。

特徴

ISO/IEC 2022 に則った形式ですが、エスケープシーケンスは用いません。GR 領域に複数の符号化文字集合を切り替えて使うには、シングルシフトコードを用います。コード値 0x8E (シングルシフト 2、SS2) または 0x8F (シングルシフト 3、SS3) です。シングルシフトは 1 文字分だけ有効な切り替えなので、ISO-2022-JP のような エスケープシーケンス による切り替えと違って状態を持つ必要がありません。つまりステートレスなコードです。

2 バイトコードと 1 バイトコードのバイト値の範囲が重なっていません。MSB が 0 のバイトは常に ASCII です。

JIS X 0213 における EUC

JIS X 0213 は附属書 3 で EUC-JIS-2004 を定義しています。内容は EUC-JIS-2004 の項を参照してください。

JIS X 0208 における EUC

EUC-JP の項を参照してください。

関連項目

- ・ EUC-JIS-2004
- ・ EUC-JP
- ・ 国際基準版・漢字用 8 ビット符号 - JIS 漢字コード規格における EUC と同等の符号