NAME

Encode::TW - Taiwan-based Chinese Encodings

SYNOPSIS

```
use Encode qw/encode decode/;
$big5 = encode("big5", $utf8); # loads Encode::TW implicitly
$utf8 = decode("big5", $big5); # ditto
```

DESCRIPTION

This module implements tradition Chinese charset encodings as used in Taiwan and Hong Kong. Encodings supported are as follows.

```
Canonical    Alias   Description
-------------------------------------------------------------------
big5-eten    /\bbig-?5$/i Big5 encoding (with ETen extensions)
     /\bbig5-?et(en)?$/i
     /\btca-?big5$/i
big5-hkscs   /\bbig5-?hk(scs)?$/i
             /\bhk(scs)?-?big5$/i
                               Big5 + Cantonese characters in Hong Kong
MacChineseTrad  Big5 + Apple Vendor Mappings
cp950                 Code Page 950
                         = Big5 + Microsoft vendor mappings
-------------------------------------------------------------------
```

To find out how to use this module in detail, see *Encode*.

NOTES

Due to size concerns, EUC-TW (Extended Unix Character), CCCII (Chinese Character Code for Information Interchange), BIG5PLUS (CMEX's Big5+) and BIG5EXT (CMEX's Big5e) are distributed separately on CPAN, under the name *Encode::HanExtra*. That module also contains extra China-based encodings.

BUGS

Since the original big5 encoding (1984) is not supported anywhere (glibc and DOS-based systems uses big5 to mean big5-eten; Microsoft uses big5 to mean cp950), a conscious decision was made to alias big5 to big5-eten, which is the de facto superset of the original big5.

The CNS11643 encoding files are not complete. For common CNS11643 manipulation, please use EUC-TW in *Encode::HanExtra*, which contains planes 1-7.

The ASCII region (0x00-0x7f) is preserved for all encodings, even though this conflicts with mappings by the Unicode Consortium. See

http://www.debian.or.jp/~kubota/unicode-symbols.html.en

to find out why it is implemented that way.

SEE ALSO

Encode