

NAME

Encode::CN - China-based Chinese Encodings

SYNOPSIS

```
use Encode qw/encode decode/;
$euc_cn = encode("euc-cn", $utf8); # loads Encode::CN implicitly
$utf8 = decode("euc-cn", $euc_cn); # ditto
```

DESCRIPTION

This module implements China-based Chinese charset encodings. Encodings supported are as follows.

Canonical Alias Description

```
-----
euc-cn /\beuc.*cn$/i EUC (Extended Unix Character)
/\bcn.*euc$/i
/\bGB[-_ ]?2312(?:\D.*$|$/i (see below)
gb2312-raw The raw (low-bit) GB2312 character map
gb12345-raw Traditional chinese counterpart to
GB2312 (raw)
iso-ir-165 GB2312 + GB6345 + GB8565 + additions
MacChineseSimp GB2312 + Apple Additions
cp936 Code Page 936, also known as GBK
(Extended GuoBiao)
hz 7-bit escaped GB2312 encoding
-----
```

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

NOTES

Due to size concerns, GB 18030 (an extension to GBK) is distributed separately on CPAN, under the name Encode::HanExtra. That module also contains extra Taiwan-based encodings.

BUGS

When you see `charset=gb2312` on mails and web pages, they really mean `euc-cn` encodings. To fix that, `gb2312` is aliased to `euc-cn`. Use `gb2312-raw` when you really mean it.

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

SEE ALSO

Encode