NAME
gb18030 - GB 18030 encoding method for Chinese text

SYNOPSIS
ENCODING "GB18030"

DESCRIPTION
The **GB18030** encoding implements GB 18030-2000, a PRC national standard for the encoding of Chinese characters. It is a superset of the older GB 2312-1980 and GBK encodings, and incorporates Unicode's Unihan Extension A completely. It also provides code space for all Unicode 3.0 code points.

Multibyte characters in the **GB18030** encoding can be one byte, two bytes, or four bytes long. There are a total of over 1.5 million code positions.

GB 11383-1981 (ASCII) characters are represented by single bytes in the range 0x00 to 0x7F.

Chinese characters are represented as either two bytes or four bytes. Characters that are represented by two bytes begin with a byte in the range 0x81-0xFE and end with a byte either in the range 0x40-0x7E or 0x80-0xFE.

Characters that are represented by four bytes begin with a byte in the range 0x81-0xFE, have a second byte in the range 0x30-0x39, a third byte in the range 0x81-0xFE and a fourth byte in the range 0x30-0x39.

SEE ALSO
euc(5), gb2312(5), gbk(5), utf8(5)

Chinese National Standard GB 18030-2000: Information Technology -- Chinese ideograms coded character set for information interchange -- Extension for the basic set, March 2000.

The Unicode Standard, Version 3.0, The Unicode Consortium, 2000.

STANDARDS
The **GB18030** encoding is believed to be compatible with GB 18030-2000.