

GB

NATIONAL STANDARD OF THE
PEOPLE'S REPUBLIC OF CHINA

ICS 35.040

L 71

GB 18030-2005

Information technology - Chinese coded character set

信息技术 中文编码字符集

Issued on: November 08, 2005

Implemented on: May 01, 2006

**Issued by: General Administration of Quality Supervision, Inspection and
Quarantine of the People's Republic of China;**

Standardization Administration of the People's Republic of China.

Table of Contents

Foreword	3
1 Scope	4
2 Normative references	4
3 Principle	4
4 Terms and definitions	5
5 Repertoire	5
6 Overall structure	6
7 Sequence of characters	8
8 Code point allocation	9
Annex A (Normative) Character table of double-byte	13
A.1 Content	13
A.2 Description	13
Annex B (Normative) Ideographic descriptors	14
Annex C (Normative) Additional Chinese characters and radicals/components	15
Annex D (Normative) Four-byte character table	16
D.1 Content	16
D.2 Description	16
Annex E (Normative) Explanation of some character codes	17

Foreword

In this Standard, the part of single-byte code, the part of double-byte code, and the part of CJK unified Chinese character extension A (i.e., 0x8139EE39-0x82358738) of the part of four-byte code are mandatory.

This Standard replaces GB 18030-2000 "Information technology - Chinese ideograms coded character set for information interchange - Extension for the basic set". Compared with GB 18030-2000, it adds the number of coded Chinese characters and supplementally stipulates the code position of some minority languages of China in this Standard. This Standard redefines the coding position of the character "𐀀". The coding architecture of this Standard remains unchanged.

Annex A, Annex B, Annex C, Annex D and Annex E of this Standard are normative.

This Standard was proposed by Ministry of Information Industry of the People's Republic of China.

This Standard shall be under the jurisdiction of China Institute of Electronic Technology Standardization.

The drafting organizations of this Standard: Electronic Industry Standardization Research Institute of the Ministry of Information Industry, Peking University Computer Technology Research Institute, Peking University Founder Group, Beijing Founder Xintiandi Information Network Technology Co., Ltd., Stone Group Corporation, China Electronics and Information Industry Development Research Institute, Chinese Academy of Sciences Institute of Software, Great Wall Software Corporation, Sitong Lifang Company, Chinasoft Corporation, Kingsoft Corporation, Lenovo Group Co., Ltd.

Main drafters of this Standard: Chen Kunqiu, Huang Jiang, Hu Wanjin, Zhang Jianguo, Chen Zhuang.

This Standard was first issued in 2000. This is the first revision.

Information technology - Chinese coded character set

1 Scope

This Standard serves as the coded character standard of the GB/T 2311 system. It specifies the hexadecimal representation of Chinese graphic characters and their binary codes used in information technology.

This Standard applies to the processing, exchange, storage, transmission, presentation, input and output of graphic character information.

2 Normative references

The provisions in following documents become the provisions of this Standard through reference in this Standard. For dated references, the subsequent amendments (excluding corrigendum) or revisions do not apply to this Standard, however, parties who reach an agreement based on this Standard are encouraged to study if the latest versions of these documents are applicable. For undated references, the latest edition of the referenced document applies.

GB/T 2311-2000, *Information technology - Character code structure and extension techniques* (idt ISO/IEC 2022:1994)

GB 2312-1980, *Code of Chinese graphic character set for information interchange - Primary set*

GB/T 11383-1989, *Information process in 8-bit code for information interchange - Structure and rules for implementation* (idt ISO 4873:1986)

GB 12345-1990, *Code of Chinese ideogram set for information interchange supplementary set*

GB 13000.1-1993, *Information technology - Universal multiple - Octet coded character set (UCS) - Part 1: Architecture and basic multilingual plane* (idt ISO/IEC 10646-1:1993)

3 Principle

This Standard is backward compatible with the internal code corresponding to the national standard GB 2312 information processing exchange code.

Regarding repertoire, this Standard supports all Chinese, Japanese and Korean (CJK)

unified Chinese characters (including CJK unified Chinese character extension A, CJK unified Chinese character extension B) characters of GB 13000 and characters of some minority languages in China.

4 Terms and definitions

The following terms and definitions apply to this Standard.

4.1 character

An element in a collection of elements used to organize, control, or represent data.

4.2 coded character

Characters and their encoded representations.

4.3 repertoire

A specified set of characters represented by a coded character set.

4.4 reserved zone

Areas reserved for future national standards to specify in this Standard.

5 Repertoire

The characters included in this Standard are coded in single-byte, double-byte or four-byte.

5.1 Part of single-byte

In this Standard, the part of single-byte includes all 128 characters from 0x00 to 0x7F of GB/T 11383-1989.

5.2 Part of double-byte

In this Standard, the contents of part of double-byte are as follows:

All CJK unified Chinese characters that are in GB 13000.1-1993. See Annex A.

21 Chinese characters that are in CJK compatible area of GB 13000.1-1993. See Annex A.

139 graphic characters that are used in Chinese Taiwan, which are included in GB 13000.1-1993 but not included in GB 2312. See Annex A.

31 other characters that are included in GB 13000.1-1993. See Annex A.

GB 18030-2005

Non-Chinese characters that are in GB 2312-1980. See Annex A.

19 vertical punctuation marks that are in GB 12345-1990. See Annex A.

10 lowercase Roman numerals that are not included in GB 2312-1980. See Annex A.

5 Chinese Pinyin letters with tones as well as *a* and *g* that are not included in GB 2312-1980. See Annex A.

The Chinese character number "〇". See Annex A.

13 ideographic descriptors. See Annex A and Annex B.

80 Chinese characters and radicals/components that are amended to GB 13000.1-1993. See Annex A and Annex C.

Double-byte coded Euro symbol. See Annex A.

5.3 Part of four-byte

The part of four-byte of this Standard includes -- in addition to the above-mentioned double-byte characters -- CJK unified Chinese character extension A, CJK unified Chinese character extension B of GB 13000, and characters of Chinese minority languages that have been coded in GB 13000. See Annex D.

6 Overall structure

In this Standard, single-byte, double-byte or four-byte characters are used to encode characters. Any byte in this Standard consists of an octet string. Any eight-bit value is represented by hexadecimal notation from 0x00 to 0xFF. In this Standard, all numbers marked with 0x are in hexadecimal. Those not marked with 0x are in decimal.

The part of single-byte adopts the encoding structure and rules of GB/T 11383-1989. Use code points 0x00 to 0x7F.

The part of double-byte adopts two octet strings to represent a character. Its first byte code bits are from 0x81 to 0xFE. The tail byte code bits are 0x40 to 0x7E and 0x80 to 0xFE respectively.

The part of four-byte adopts 0x30 to 0x39 not used in GB/T 11383-1989 as the suffix to expand the double-byte code. The four-byte code extended in this way ranges from 0x81308130 to 0xFE39FE39. The first byte of a four-byte character codes in the range 0x81 to 0xFE. The second byte code ranges from 0x30 to 0x39. The third byte code ranges from 0x81 to 0xFE. The fourth byte code ranges from 0x30 to 0x39. That is,

0x81308130 to 0x81308139;

GB 18030-2005

corresponding characters of basic multilingual plane in GB 13000. The remaining code points are reserved.

There is a total of 12600 code points from 0x85308130 to 0x8539FE39. It is the reserved zone, reserved for future character extensions.

There is a total of 126000 code points from 0x86308130 to 0x8F39FE39. It is the reserved zone, reserved for the extension of Chinese characters in the future.

There is a total of 1058400 code points from 0x90308130 to 0xE339FE39. It is used to correspond to 16 auxiliary planes of GB 13000. The sequence of characters is completely in accordance with the corresponding code point sequence of the 16 auxiliary planes of GB 13000. The remaining code points are reserved.

There is a total of 315000 code points from 0xE4308130 to 0xFC39FE39. It is the reserved zone, reserved for future standard extensions.

There is a total of 25200 code points from 0xFD308130 to 0xFE39FE39. It is the user-defined zone.

See Annex D.

8 Code point allocation

8.1 Code point allocation for part of single-byte

In this Standard, refer to GB/T 11383-1989 for the code point allocation of the part of single-byte. See Figure 2.

This is an excerpt of the PDF (Some pages are marked off intentionally)

Full-copy PDF can be purchased from 1 of 2 websites:

1. <https://www.ChineseStandard.us>

- SEARCH the standard ID, such as GB 4943.1-2022.
- Select your country (currency), for example: USA (USD); Germany (Euro).
- Full-copy of PDF (text-editable, true-PDF) can be downloaded in 9 seconds.
- Tax invoice can be downloaded in 9 seconds.
- Receiving emails in 9 seconds (with download links).

2. <https://www.ChineseStandard.net>

- SEARCH the standard ID, such as GB 4943.1-2022.
- Add to cart. Only accept USD (other currencies - <https://www.ChineseStandard.us>).
- Full-copy of PDF (text-editable, true-PDF) can be downloaded in 9 seconds.
- Receiving emails in 9 seconds (with PDFs attached, invoice and download links).

Translated by: Field Test Asia Pte. Ltd. (Incorporated & taxed in Singapore. Tax ID: 201302277C)

About Us (Goodwill, Policies, Fair Trading...): <https://www.chinesestandard.net/AboutUs.aspx>

Contact: Wayne Zheng, Sales@ChineseStandard.net

Linkin: <https://www.linkedin.com/in/waynezhengwenrui/>

----- The End -----