

ICS 35.040
CCS L71



**NATIONAL STANDARD OF THE
PEOPLE'S REPUBLIC OF CHINA**

GB 18030-2022

Information technology - Chinese coded character set

信息技术 中文编码字符集

Issued on: July 19, 2022

Implemented on: August 1, 2023

**Issued by: State Administration for Market Regulation;
Standardization Administration of China**

Table of Contents

1 Scope
2 Normative references
3 Terms and definitions
4 Repertoire
5 Overall structure
6 Sequence of characters
7 Code point allocation
8 Explanation of some characters and codes
9 Implementation level
9.2 Implementation level
9.3 Implementation level
9.4 Implementation level
Annex A
Annex C

1 Scope

GB 18030-2022 is the Chinese national standard on information technology - chinese coded character set, in the field of information technology. It carries no /T suffix, which in the Chinese system means compliance is mandatory: a product or a practice within its scope has to meet it to be lawfully made, sold or carried out in China. It was issued on 19 July 2022 by the State Administration for Market Regulation; Standardization Administration of China, and has been in force since 1 August 2023. Classification: ICS 35.040, CCS L71. This page is published from the official record of the standard held by the Chinese standards administration: the identification, the dates, the classification and the issuing body are taken from there. The clause text, the tables and the numeric limits are in the document itself, which is delivered complete in English translation.

This document specifies the hexadecimal representation of Chinese graphic characters and their binary codes used in information technology. This document applies to the processing, exchange, storage, transmission, presentation, input and output of Chinese and other graphic character information. This document is applicable to technical products with information processing and exchange functions of Chinese and other text and graphic characters, including but not limited to the software products represented by input methods, optical character recognition (OCR), editing and proofreading, machine translation, speech

synthesis, text transcription, intelligent writing, etc., as well as the hardware products represented by computers, communication terminal equipment, e-book readers, learning machines, etc.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

GB/T 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

GB/T 13000, Information technology - Universal multiple - Octet coded character set (UCS)

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

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

3.2 coded character Character (3.1) and its coded representation.

3.3 private use area An area that can be specified by the user of a product conforming to this document.

3.4 repertoire A specified set of characters (3.1) represented by a coded character (3.2) set.

3.5 reserved zone Areas reserved for future specified by this document.

4 Repertoire

4.1 Overview The characters included in this document are coded in single-byte, double-byte or four- byte.

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

4.3 Part of double-byte The part of double-byte includes all graphic characters in GB/T 2312-1980, CJK unified Chinese characters and some graphic characters in GB/T 13000. The characters in the part of double-byte are in accordance with the provisions in Annex A. Among them, the graphics, code positions and functions of ideographic descriptors shall comply with the provisions of Annex B. NOTE. GB/T 13000 uniformly

encodes Chinese characters used in China, Japan, South Korea, Vietnam and other countries and regions. Chinese characters with unique abstract glyphs are assigned a separate code position. Chinese characters with different sources but the same abstract glyphs are given a common code position. The encoded Chinese characters are called CJK unified Chinese characters (CJK Unified Ideographs), where CJK means China, Japan, and Korea.

4.4 Part of four-byte The part of four-byte includes 66 CJK unified Chinese characters (9FA6~9FEF, excluding 9FB4~9FBB) in GB/T 13000 other than the above-mentioned double-byte characters, CJK unified Chinese character extension A, CJK unified Chinese character extension B, CJK unified Chinese character extension C, CJK unified Chinese character extension D, CJK unified Chinese character extension E, CJK unified Chinese character extension F and the characters of ethnic minorities that have been coded in GB/T 13000. The characters in the part of four-byte follow the provisions of Annex C.

5 Overall structure

In the text, all numbers marked with 0x are in hexadecimal. Those not marked with 0x are in decimal. All coded representations in the appendix are expressed in hexadecimal. All other numbers are expressed in decimal. The part of single-byte adopts the encoding structure of GB/T 11383-1989. Use code points 0x00~0x7F.

6 Sequence of characters

6.1 Sequence of characters in part of single-byte All characters in the part of single-byte are arranged in the order of the corresponding characters in GB/T 11383-1989.

6.2 Sequence of characters in part of double-byte See Annex A for the sequence of characters in the part of double-byte.

6.3 Sequence of characters in part of four-byte There is a total of 50400 code points from 0x81308130 to 0x8439FE

39. The characters corresponding to all basic multilingual plane of GB/T 13000 not included in the part of double-byte shall be arranged in the order of the corresponding characters of basic multilingual plane in GB/T 13000. A total of 1058400 code points from 0x90308130 to 0xE339FE39 are used for the 16 auxiliary planes corresponding to GB/T 13000. The sequence of character arrangement is completely in accordance with the corresponding code point sequence of the 16 auxiliary planes of GB/T 13000. The sequence of characters in the part of four-byte shall comply with Annex C.

7 Code point allocation

7.1 Code point allocation for part of single-byte The code points of the part of single-byte

are allocated according to the rules of GB/T 11383-1989. See Figure 2 for the allocation of single-byte code points. Figure 2 -- Code point map for zone of single byte

7.2 Code point allocation for part of double-byte The code point arrangement of the part of double-byte is divided into two parts. 0x8140~0xFE7E and 0x8180~0xFEFE, a total of 23940 code points. See Figure 3 and Table 2 for the allocation of double-byte code points.

8 Explanation of some characters and codes

Compared with GB 18030-2005, the glyphs at some code positions and/or the corresponding GB/T 13000 code positions have been adjusted in this document (see Annex D).

9 Implementation level

9.1 General This document specifies three implementation levels. System software products that meet the corresponding implementation level shall provide input and output functions for all characters within the corresponding implementation level.

9.2 Implementation level

1 Implementation level 1 supports CJK unified Chinese characters (i.e., 0x82358F33~0x82359636) and CJK unified Chinese character extension A (i.e., 0x8139EE39~0x82358738) of the single-byte coded part, double-byte coded part and four-byte coded part of this document. Any product to which this document applies shall meet the requirements for implementation level 1. NOTE. According to the needs of software applications, implementation level 1 can also choose to support any one or more non-Chinese characters listed in Table 3.

Annex A

(normative) Character table of double-byte A.1 Content This table gives the glyphs and codes of double-byte coded characters. At the same time, the GB/T 13000 code position corresponding to this character is given. A.2 Description

Annex C

(normative) Character table of four-byte C.1 Contents This table gives the glyphs and codes of Chinese characters and some minority languages. At the same time, the GB/T 13000 code position corresponding to this character is also given. The fonts of other parts are omitted. C.2 Description