



# 中华人民共和国国家标准

GB/T 13000—2025

代替 GB/T 13000—2010

## 信息技术 通用编码字符集(UCS)

Information technology—Universal coded character set (UCS)

(ISO/IEC 10646:2020, MOD)

2025-01-24 发布

2025-08-01 实施

国家市场监督管理总局  
国家标准化管理委员会 发布

## 目 次

前言 .....	VII
1 范围 .....	1
2 规范性引用文件 .....	1
3 术语和定义 .....	2
4 符合性 .....	8
4.1 通用要求 .....	8
4.2 信息交换的符合性 .....	8
4.3 设备的符合性 .....	8
5 电子数据附件 .....	9
6 UCS 的总体结构 .....	10
7 基本结构与命名法 .....	10
7.1 结构 .....	10
7.2 字符的编码 .....	11
7.3 码位类型 .....	12
7.4 字符的命名 .....	13
7.5 字符别名 .....	14
7.6 码位的短标识符(UID) .....	14
7.7 UCS 序列标识符 .....	15
7.8 八位序列标识符 .....	15
8 UCS 的修订与更新 .....	15
9 子集 .....	15
9.1 概述 .....	15
9.2 有限子集 .....	15
9.3 选择子集 .....	15
10 UCS 编码形式 .....	16
10.1 概述 .....	16
10.2 UTF-8 .....	16
10.3 UTF-16 .....	17
10.4 UTF-32 .....	17
11 UCS 编码方案 .....	18
11.1 概述 .....	18
11.2 UTF-8 .....	18
11.3 UTF-16BE .....	18

11.4	UTF-16LE	18
11.5	UTF-16	18
11.6	UTF-32BE	19
11.7	UTF-32LE	19
11.8	UTF-32	19
12	控制功能与 UCS 的联合使用	19
13	标识特征的声明	20
13.1	标识的目的与缘由	20
13.2	UCS 编码方案的标识	20
13.3	图形字符子集的标识	21
13.4	控制功能集的标识	21
13.5	ISO/IEC 2022 编码体系的标识	21
14	代码表与字符名称清单的结构	22
15	块和汇集的名称	22
15.1	块名称	22
15.2	汇集名称	22
16	双向语境中的镜像字符	23
16.1	镜像字符	23
16.2	双向语境的方向性	23
17	特殊字符	23
17.1	概述	23
17.2	间隔字符	23
17.3	货币符号	24
17.4	格式字符	24
17.5	表意文字描述符	26
17.6	变体选择符与变体序列	27
18	字符的显现形式	29
19	兼容字符	29
20	字符的顺序	29
21	组合用字符	29
21.1	组合用字符的顺序	29
21.2	组合类别与正则顺序	30
21.3	代码表中的显示	30
21.4	替代编码表示	30
21.5	多重组合用字符	30
21.6	含组合用字符的汇集	31
21.7	组合用字素连接符	31

22	规范化形式 .....	31
23	个别文字及符号字汇的特性 .....	32
23.1	谚文(朝鲜文)音节复合法 .....	32
23.2	印度和南亚一些国家的文字特性 .....	32
23.3	拜占庭派音乐符号 .....	33
23.4	图形符号的字源参考 .....	33
24	CJK 汉字的字源参考 .....	33
24.1	字源参考文件清单 .....	33
24.2	CJK 汉字的字源参考文件 .....	37
24.3	CJK 统一汉字的字源参考信息表示方式 .....	40
24.4	CJK 兼容汉字的字源参考信息表示方式 .....	42
25	西夏字的字源参考 .....	43
25.1	字源参考清单 .....	43
25.2	西夏字的字源参考文件 .....	43
25.3	西夏字的字源参考信息表示方式 .....	44
26	女书字符的字源参考 .....	45
26.1	字源参考清单 .....	45
26.2	女书字符的字源参考文件 .....	45
27	字符名称及注释 .....	46
27.1	实体名称 .....	46
27.2	名称构成 .....	46
27.3	单一名称 .....	47
27.4	名称不变性 .....	47
27.5	名称唯一性 .....	47
27.6	CJK 汉字的字符名称 .....	48
27.7	西夏字的字符名称 .....	49
27.8	女书的字符名称 .....	49
27.9	契丹小字的字符名称 .....	49
27.10	谚文(朝鲜文)音节的字符名称 .....	49
28	命名的 UCS 序列标识符 .....	51
29	基本多文种平面(BMP)的结构 .....	52
30	用于文字和符号编码的辅助多文种平面(SMP)的结构 .....	54
31	辅助表意文字平面(SIP)的结构 .....	57
32	第三表意文字平面(TIP)的结构 .....	57
33	辅助特殊用途平面(SSP)的结构 .....	58
34	代码表和字符名称清单 .....	58
34.1	概述 .....	58

34.2	代码表	58
34.3	字符名称清单	59
34.4	标准化变体序列汇总	60
34.5	代码表和字符名称清单	60
附录 A (资料性)	用于标识符的字符	72
附录 B (资料性)	对字符字汇的外部引用	73
B.1	字符字汇及其编码的引用方法	73
B.2	ASN.1 字符抽象语法的标识	73
B.3	ASN.1 字符传送语法的标识	74
附录 C (资料性)	八位值表示的标记法	75
附录 D (资料性)	对具有内部存储器的接收、始发组合设备的建议	76
附录 E (规范性)	用于子集的图形字符的汇集	77
E.1	已编码图形字符的汇集	77
E.2	块名称清单	91
E.3	整个 UCS 的固定汇集(Unicode 汇集除外)	99
E.4	CJK 汇集	104
E.5	其他汇集	106
E.6	Unicode 汇集	113
附录 F (资料性)	按字母排序的字符名称清单	114
附录 G (资料性)	格式字符	115
G.1	概述	115
G.2	通用格式字符	115
G.3	适用于特定文字的格式字符	117
G.4	行间注释字符	118
G.5	反向格式字符	118
G.6	速记格式字符	118
G.7	隐形数学运算符	119
G.8	西方音乐符号	119
G.9	使用标签字符的语言标记法	120
附录 H (资料性)	表意文字描述符	122
H.1	概述	122
H.2	表意文字描述序列的语法	122
H.3	表意文字描述符的个体定义	122
附录 I (资料性)	CJK 汉字的认同和排序规则	125
I.1	概述	125
I.2	认同程序	125
I.3	排序程序	128

I.4 源字集分离示例 .....	129
I.5 不认同示例 .....	134
I.6 对“CJK 汉字的认同和排序规则”的补充说明 .....	135
附录 J (资料性) 字符命名导则 .....	137
附录 K (资料性) 谚文(朝鲜文)音节名称 .....	140
附录 L (资料性) 部分中国文字和符号的补充说明 .....	141
L.1 概述 .....	141
L.2 部分中国文字和符号的名称及在本文件中的英文译名 .....	141
L.3 代码表中关于部分中国文字的说明 .....	142
附录 M (资料性) CJK 统一汉字附加说明 .....	144
附录 N (规范性) 代码表和字符名称清单 .....	147
参考文献 .....	2926

## 前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第 1 部分：标准化文件的结构和起草规则》的规定起草。

本文件代替 GB/T 13000—2010《信息技术 通用多八位编码字符集(UCS)》，与 GB/T 13000—2010 相比，除结构调整和编辑性改动外，主要技术变化如下：

- a) 增加了范围中对于第三表意文字平面(TIP)、UCS 的三种编码形式和七种编码方案的规定(见第 1 章)；
- b) 更改了术语和定义中的部分内容(见第 3 章,2010 年版的第 4 章)；
- c) 更改了信息交换的符合性(见 4.2,2010 年版的 2.2)和设备的符合性(见 4.3,2010 年版的 2.3)的规定；
- d) 增加了“电子数据附件”一章(见第 5 章)；
- e) 更改了 UCS 代码空间的范围,在 UCS 的总体结构中增加了第三表意文字平面(TIP)(见第 6 章、7.1,2010 年版的第 5 章、6.1)；
- f) 更改了字符编码的表现格式(见 7.2,2010 年版的 6.2)，删除了八位的次序的规定(见 2010 年版的 6.3)，增加了码位类型的规定(见 7.3)，更改了码位的短标识符(UID)的格式(见 7.6,2010 年版的 6.5)，增加了八位序列标识符的规定(见 7.8)；
- g) 增加了“UCS 编码形式”一章(见第 10 章)；
- h) 增加了“UCS 编码方案”一章(见第 11 章)；
- i) 删除了“实现的级别”一章(见 2010 年版的第 14 章)；
- j) 增加了 UCS 编码方案的标识(见 13.2)，删除了带实现级别的 UCS 编码表示形式的标识(见 2010 年版的 16.2)；
- k) 更改了镜像字符的规定(见第 16 章,2010 年版的第 19 章、附录 E)；
- l) 增加了表意文字描述符(见 17.5)、变体选择符和变体选择序列的规定(见 17.6)；
- m) 增加了组合类别及正则顺序(见 21.2)、组合用字素连接符的规定(见 21.7)；
- n) 增加了印度和南亚一些国家文字的唯一拼写规则(见 23.2)和图形符号的字源参考(见 23.4)；
- o) 增加了 CJK 汉字字源标准分类(见 24.1)，更改了字源参考清单(见 24.1.2010 年版的 27.1)，增加了 CJK 汉字字源参考文件(见 24.2)、CJK 统一汉字的字源参考信息表示方式(见 24.3)、CJK 兼容汉字的字源参考信息表示方式(见 24.4)；
- p) 增加了“西夏字的字源参考”“女书字符的字源参考”“字符名称及注释”和“命名的 UCS 序列标识符”四章(见第 25 章、第 26 章、第 27 章、第 28 章)；
- q) 增加了新文种(见第 29 章、第 30 章、第 31 章、第 33 章)；
- r) 增加了“第三表意文字平面(TIP)的结构”一章(见第 32 章)；
- s) 更改了代码表、字符名称清单和字符来源(见第 34 章,2010 年版的第 33 章)；
- t) 更改了用于子集的图形字符汇集的清单(见附录 E,2010 年版的附录 A)。

本文件修改采用 ISO/IEC 10646:2020《信息技术 通用编码字符集(UCS)》和 ISO/IEC 10646:2020/Amd 1:2023。

本文件与 ISO/IEC 10646:2020 和 ISO/IEC 10646:2020/Amd 1:2023 相比做了下述结构调整：

- 为避免出现悬置段，增加了 27.5.1、G.1、I.1、I.2.3.1 和 I.2.4.1，后面各章条号依次顺延；
- 34.1 的注 2 对应 ISO/IEC 10646:2020 中的附录 M；

- 34.2 的注 4 对应 ISO/IEC 10646:2020 中的附录 Q；
- 附录 A 对应 ISO/IEC 10646:2020 中的附录 U；
- 附录 B 对应 ISO/IEC 10646:2020 中的附录 N；
- 附录 C 对应 ISO/IEC 10646:2020 中的附录 K；
- 附录 D 对应 ISO/IEC 10646:2020 中的附录 J；
- 附录 E 对应 ISO/IEC 10646:2020 中的附录 A；
- 附录 F 对应 ISO/IEC 10646:2020 中的附录 G；
- 附录 G 对应 ISO/IEC 10646:2020 中的附录 F；
- 附录 H 对应 ISO/IEC 10646:2020 中的附录 I,其中增加了 H.2.1 和 H.2.2；
- 附录 I 对应 ISO/IEC 10646:2020 中的附录 S；
- 附录 J 对应 ISO/IEC 10646:2020 中的附录 L；
- 附录 K 对应 ISO/IEC 10646:2020 中的附录 R；
- 附录 M 对应 ISO/IEC 10646:2020 中的附录 P；
- 附录 N 对应 ISO/IEC 10646:2020 中的 34.5；
- 删除了 ISO/IEC 10646:2020 的附录 B、附录 C、附录 D、附录 E、附录 H 和附录 T。

本文件与 ISO/IEC 10646:2020 和 ISO/IEC 10646:2020/Amd 1:2023 的技术差异及其原因如下：

- 将适用于非英文版本的实体名称构成规则补充内容由注移入了正文,以符合国家标准起草规则；
- 删除了关于已废止的文件或标准的内容,并用脚注加以说明,以符合国家标准起草规则；
- 更改了国际标准中引用的不准确的标准编号,并用脚注加以说明,以符合国家标准起草规则；
- 在 TIP 中的块的清单中增加了“CJK UNIFIED IDEOGRAPHS EXTENSION H”,以与 ISO/IEC 10646:2020/Amd 1:2023 的其他新增内容一致。

本文件做了下列编辑性改动：

- 更改了术语“专用平面”的注,补充了其他常用译名；
- 增加了关于专用字符使用说明的注；
- 增加了关于 CJK 汉字认同和排序的内容；
- 增加了关于字符来源清单的注,其内容来自 ISO/IEC 10646:2020 的附录 M；
- 增加了关于朝鲜文音节代码映射表查询方式的注,其内容来自 ISO/IEC 10646:2020 的附录 Q；
- 增加了块名称的中文译名；
- 更改了 ASN.1 字符传送语法的标识相关内容中引用文件的附录编号；
- 增加了 ISO/IEC 10646:2020/Amd 1:2023 遗漏的不在字符名称清单文件中列出字符名称的块(CJK UNIFIED IDEOGRAPHS EXTENSION H)；
- 增加了附录 L(资料性)“部分中国文字和符号的补充说明”；
- 增加了参考文献。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由全国信息技术标准化技术委员会(SAC/TC 28)提出并归口。

本文件起草单位:中国电子技术标准化研究院、教育部语言文字应用研究所、中国科学院软件研究所、北京北大方正电子有限公司。

本文件主要起草人:陈壮、陈晓研、何正安、黄姗姗、王晓明、吴健、陈恳。

本文件及其所代替文件的历次版本发布情况为：

- 1993 年首次发布为 GB 13000.1—1993《信息技术 通用多八位编码字符集(UCS) 第一部分:体系结构与基本多文种平面》；



- 2010年第一次修订为 GB/T 13000—2010《信息技术 通用多八位编码字符集(UCS)》；
- 本次为第二次修订。

# 信息技术 通用编码字符集(UCS)

## 1 范围

本文件：

- 规定了 UCS 的体系结构；
- 定义了 UCS 中使用的术语；
- 描述了 UCS 代码空间的总体结构；
- 规定了 UCS 已分配平面：UCS 基本多文种平面(BMP)、辅助多文种平面(SMP)、辅助表意文字平面(SIP)、第三表意文字平面(TIP)以及辅助特殊用途平面(SSP)；
- 定义了用于全球各文种书写形式的图形字符集；
- 规定了 UCS 的 BMP、SMP、SIP、TIP、SSP 的图形字符与格式字符的名称及编码表示；
- 规定了控制字符与专用字符的编码表示；
- 规定了 UCS 的三种编码形式：UTF-8、UTF-16 和 UTF-32；
- 规定了 UCS 的七种编码方案：UTF-8、UTF-16、UTF-16BE、UTF-16LE、UTF-32、UTF-32BE 和 UTF-32LE；
- 规定了未来增补编码字符的管理办法。

本文件适用于具备全球各文种图形字符信息化处理及交换功能的技术类产品。

注：本文件对其编码的字符是否适合用作编程语言的标识符不做规定。附录 A 给出了关于适合用作标识符的字符的参考文件。

## 2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本(包括所有的修改单)适用于本文件。

ISO/IEC 2022 信息技术 字符代码结构与扩充技术(Information technology—Character code structure and extension techniques)

注：GB/T 2311—2000 信息技术 字符代码结构与扩充技术(ISO/IEC 2022:1994, IDT)

ISO/IEC 6429 信息技术 编码字符集的控制功能(Information technology—Control functions for coded character sets)

注：GB/T 5261—1994 信息处理 七位和八位编码字符集用的控制功能(ISO 6429:1988, IDT)

Unicode Standard Annex, UAX #9, The Unicode Bidirectional Algorithm:

<https://www.unicode.org/reports/tr9/tr9-46.html>

Unicode Standard Annex, UAX #15, Unicode Normalization Forms:

<https://www.unicode.org/reports/tr15/tr15-53.html>

Unicode Technical Standard, UTS #37, Ideographic Variation Database:

<http://www.unicode.org/reports/tr37/tr37-12.html>

Unicode Standard Version 15.0, Chapter 4, Character Properties: