



中华人民共和国国家标准

GB 13000—2010/ISO/IEC 10646:2003
代替 GB 13000.1—1993

信息技术 通用多八位编码字符集(UCS)

Information technology—Universal multiple-octet coded character set (UCS)

(ISO/IEC 10646:2003, IDT)

自 2017 年 3 月 23 日起,本标准转为推荐性
标准,编号改为 **GB/T 13000—2010**。

2011-01-10 发布

2011-11-01 实施

中华人民共和国国家质量监督检验检疫总局
中国国家标准化管理委员会 发布

中 华 人 民 共 和 国
国 家 标 准
信息技术 通用多八位编码字符集(UCS)
GB 13000—2010/ISO/IEC 10646:2003

*

中国标准出版社出版发行
北京复兴门外三里河北街16号
邮政编码:100045

网址:www.gb168.cn

服务热线:010-68522006

2011年7月第一版

*

书号:155066·1-42160

版权专有 侵权必究

目 次

前言	III
引言	IV
1 范围	1
2 符合性	1
3 规范性引用文件	2
4 术语和定义	2
5 UCS 的总体结构	6
6 基本结构及术语	7
7 UCS 的总体要求	10
8 基本多文种平面	10
9 辅助平面	11
10 专用平面和专用区	11
11 UCS 的修订与更新	11
12 子集	12
13 UCS 的编码表示形式	12
14 实现的级别	12
15 控制功能与 UCS 的联合使用	13
16 标识特征的声明	13
17 代码表与字符名称清单的结构	14
18 块名称	15
19 双向语境中的字符	15
20 特殊字符	16
21 字符的显现形式	22
22 兼容字符	22
23 字符的顺序	22
24 规范化形式	22
25 组合用字符	23
26 个别文字的特性	24
27 CJK 汉字的字源参考	25
28 字符名称及注释	28
29 基本多文种平面的结构	30
30 用于文字和符号编码的辅助多文种平面结构	32
31 辅助表意文字平面的结构	33
32 辅助特殊用途平面的结构	34
33 代码表和字符名称清单	34
附录 A (规范性附录) 用于子集的图形字符的汇集	1353
附录 B (规范性附录) 组合字符清单	1373
附录 C (规范性附录) 00 组 16 个平面的转换格式(UTF-16)	1388

附录 D (规范性附录)	UCS 转换格式 8(UTF-8)	1392
附录 E (资料性附录)	双向语境中的镜像字符	1397
附录 F (资料性附录)	替代格式字符	1403
附录 G (资料性附录)	按字母排序的字符名称清单	1410
附录 H (资料性附录)	使用“标记”标识 UCS	1411
附录 J (资料性附录)	对具有内部存储器的接收/始发组合设备的建议	1412
附录 K (资料性附录)	八位值表示的标记法	1413
附录 L (资料性附录)	字符命名导则	1414
附录 M (资料性附录)	字符来源	1417
附录 N (资料性附录)	对字符字汇的外部参考	1422
附录 P (资料性附录)	字符的附加信息	1424
附录 Q (资料性附录)	朝鲜文代码映射表	1428
附录 R (资料性附录)	朝鲜文名称	1429
附录 S (资料性附录)	CJK 汉字的认同和排序规则	1454
附录 T (资料性附录)	使用标签字符的语言标记法	1467
附录 U (资料性附录)	音乐符号的使用	1469

前 言

本标准的体系结构与基本多文种平面部分是强制性的,其余为推荐性的。

本标准等同采用国际标准 ISO/IEC 10646:2003《信息技术 通用多八位编码字符集(UCS)》(英文版)。为便于读者理解,本标准在第 1 章增加了脚注,在附录 S(资料性附录)的最后增加了“S. 4 对‘CJK 汉字认同和排序规则’的补充说明”。

本标准代替 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》。

本次修订对 1993 年版的主要变动如下:

- a) 朝鲜文及其补充从基本多文种平面的 3400~4DFF 移至基本多文种平面的 AC00~D7FF(原 O 区,即第一版的保留区),空出的代码位置分配给新增的 CJK 统一汉字扩充 A;
- b) 增加收录多种我国少数民族文字及其他文字、字符,如藏文、蒙古文、彝文等;
- c) 新增用于 UTF-16 的 S 区(代理区),代码位置是基本多文种平面的 D800~DFFF(原 O 区,即第一版的保留区),并有专门的附录对其进行说明;
- d) 增加了辅助平面,包括 00 组 01 平面(文字和符号辅助多文种平面)、00 组 02 平面(辅助表意文字平面,用于 CJK 统一汉字扩充 B 和 CJK 兼容汉字补充)和 00 组 0E 平面(辅助特殊用途平面)。

本标准的附录 A、附录 B、附录 C 和附录 D 是规范性附录,附录 E、附录 F、附录 G、附录 H、附录 J、附录 K、附录 L、附录 M、附录 N、附录 P、附录 Q、附录 R、附录 S、附录 T 和附录 U 是资料性附录。

本标准由中华人民共和国工业和信息化部提出。

本标准由全国信息技术标准化技术委员会(SAC/TC 28)归口。

本标准起草单位:中国电子信息产业发展研究院、中国电子技术标准化研究所、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。

本标准主要起草人:张轴材、陈壮、王晓明、吴健、尹江红、何正安。

本标准于 1993 年首次发布,本次为第一次修订。

根据中华人民共和国国家标准公告(2017 年第 7 号)和强制性标准整合精简结论,本标准自 2017 年 3 月 23 日起,转为推荐性标准,不再强制执行。

引 言

本标准规定了通用多八位编码字符集(UCS)。它适用于世界上各种语言(文字)的书面形式以及附加符号的表示、传输、交换、处理、存储、输入及显现。

通过对多文种文本的编码的一致性的定义,本标准使得数据的国际交换成为可能。信息技术产业获得了数据的稳定性,更强的全球可互操作性和数据可交换性。转化为本国家标准的国际标准 ISO/IEC 10646 已经在新的互联网协议中广泛采用,并被当今的操作系统和计算机语言所实现。本标准容纳了 95 000 多个世界上多种文字的字符。

本标准包含了一些电子化资料,它们适用于使用机读格式数据的用户。这些资料由下列可打印文件组成:

- CJKU_SR.txt
- CJKC_SR.txt
- Allnames.txt
- HangulX.txt
- HangulSy.txt

用户可向本标准归口单位(中国电子技术标准化研究所)索取上述电子化资料。

地址:北京市东城区安定门东大街 1 号(北京市 1101 信箱)

邮编:100007

电话:84043004

信息技术 通用多八位编码字符集(UCS)

1 范围

本标准规定了通用多八位编码字符集(UCS)。

本标准适用于世界上各种语言(文字)的书面形式以及附加符号的表示、传输、交换、处理、存储、输入及显现。

本标准：

- 规定了 GB 13000 的体系结构；
- 定义了 GB 13000 中使用的术语；
- 描述了本编码字符集的总体结构；
- 规定了 UCS 的基本多文种平面(BMP)；
- 规定了 UCS 的若干个辅助平面：辅助多文种平面(SMP)、辅助表意文字平面(SIP)，以及辅助特殊用途平面(SSP)；
- 定义了一个图形字符集，用于世界各种语言的手写和书面形式；
- 规定了 BMP、SMP、SIP、SSP 的图形字符的名称及编码表示；
- 规定了 UCS 的肆八位(32 位)正则形式：UCS-4；
- 规定了 UCS 的双八位(16 位)BMP 形式：UCS-2；
- 规定了控制功能的编码表示；
- 规定了未来对编码字符集进行增补的管理办法。

UCS 是一种与 GB/T 2311 规定的编码体系不同的编码体系。本标准 16.2 规定了从 GB/T 2311 中指明 UCS 的方法。

本标准中任意一个图形字符，无论是在 BMP 平面还是在辅助平面，只分配唯一的一个码位¹⁾。

注：Unicode Standard 的 4.0 版包括了与本标准等同的字符集、名称和编码表示。为了便于实现，它还提供了字符属性、处理算法以及定义的细节。

2 符合性

2.1 总则

专用字符一旦按本标准规定的方法使用，这些专用字符本身就不再受下列符合性要求的约束。

2.2 信息交换的符合性

如在编码信息内用于交换的编码字符数据元素符合以下条件，则符合本标准：

- a) 在该编码字符数据元素内的全部图形字符的编码表示都符合第 6 章和第 7 章，及选自第 13 章或附录 C 或附录 D 的一种已标识的形式，并且还符合选自第 14 章的一种已标识的实现级别；
- b) 在该编码字符数据元素内所表示的全部图形字符都来源于一个已标识的子集(见第 12 章)；
- c) 在该编码字符数据元素内的全部控制功能的编码表示都符合第 15 章。

符合性声明必须标识出所采用的形式、所采用的实现级别以及所采用的以汇集清单和(或)字符清

1) 实际上，上述“一字一码”的原则在本标准的下列情形中，作了特别处理：

- a) 不同文种的同形字符是分别编码的，比如拉丁 A，希里尔文 A 和希腊文 A，就是分别编码的；
- b) CJK 同一源字集中的同形字符，可能在 CJK 统一汉字编码区和 CJK 兼容汉字编码区中分别编码；
- c) CJK 统一汉字中可做部首的汉字，可能在 Kangxi Radical 或 CJK Radical Supplement 中有同形的编码部首。