XML作业

**1.GB 2312：**

**来历：GB 2312**是[中华人民共和国国家标准](https://zh.wikipedia.org/wiki/%E4%B8%AD%E5%8D%8E%E4%BA%BA%E6%B0%91%E5%85%B1%E5%92%8C%E5%9B%BD%E5%9B%BD%E5%AE%B6%E6%A0%87%E5%87%86" \o "中华人民共和国国家标准)[简体中文](https://zh.wikipedia.org/wiki/%E7%AE%80%E4%BD%93%E4%B8%AD%E6%96%87" \o "简体中文)[字符集](https://zh.wikipedia.org/wiki/%E5%AD%97%E7%AC%A6%E9%9B%86" \o "字符集)，全称《**信息交换用汉字编码字符集·基本集**》，又称**[GB0](https://zh.wikipedia.org/wiki/%E5%9B%BD%E5%AE%B6%E6%A0%87%E5%87%86%E4%BB%A3%E7%A0%81" \o "国家标准代码)**，由[中国国家标准总局](https://zh.wikipedia.org/w/index.php?title=%E4%B8%AD%E5%9B%BD%E5%9B%BD%E5%AE%B6%E6%A0%87%E5%87%86%E6%80%BB%E5%B1%80&action=edit&redlink=1" \o "中国国家标准总局（页面不存在）)发布，1981年5月1日实施。GB 2312编码通行于中国大陆；[新加坡](https://zh.wikipedia.org/wiki/%E6%96%B0%E5%8A%A0%E5%9D%A1" \o "新加坡)等地也采用此编码。中国大陆几乎所有的中文系统和国际化的软件都支持GB 2312。

**产生原因：**信息交换用汉字编码字符集和[汉字输入](https://baike.baidu.com/item/%E6%B1%89%E5%AD%97%E8%BE%93%E5%85%A5" \t "https://baike.baidu.com/item/%E4%BF%A1%E6%81%AF%E4%BA%A4%E6%8D%A2%E7%94%A8%E6%B1%89%E5%AD%97%E7%BC%96%E7%A0%81%E5%AD%97%E7%AC%A6%E9%9B%86/_blank)编码之间的关系是，根据不同的汉字输入方法，通过必要的设备向计算机输入汉字的编码，计算机接收之后，先转换成信息交换用汉字编码字符，这时计算机就可以识别并进行处理；汉字输出是先把机内码转成汉字编码，再发送到[输出设备](https://baike.baidu.com/item/%E8%BE%93%E5%87%BA%E8%AE%BE%E5%A4%87" \t "https://baike.baidu.com/item/%E4%BF%A1%E6%81%AF%E4%BA%A4%E6%8D%A2%E7%94%A8%E6%B1%89%E5%AD%97%E7%BC%96%E7%A0%81%E5%AD%97%E7%AC%A6%E9%9B%86/_blank)。国家标准《信息交换用汉字编码字符集·基本集》已于1981年5月发布实施。《基本集》规定了汉字信息交换用的基本图形字符及其二进制编码,收汉字6763个。它适用于一般汉字处理、汉字通信等系统之间的信息交换。随着我国汉字信息处理技术的发展,计算机的应用范围不断扩大,使用汉字字数较多的部门迫切需要在《基本集》的基础上继续制定信息交换用汉字编码字符集各辅助集的国家标准。

**解决的问题：**GB 2312标准共收录6763个汉字，其中一级汉字3755个，二级汉字3008个；同时，GB 2312收录了包括[拉丁字母](https://baike.baidu.com/item/%E6%8B%89%E4%B8%81%E5%AD%97%E6%AF%8D" \t "https://baike.baidu.com/item/%E4%BF%A1%E6%81%AF%E4%BA%A4%E6%8D%A2%E7%94%A8%E6%B1%89%E5%AD%97%E7%BC%96%E7%A0%81%E5%AD%97%E7%AC%A6%E9%9B%86/_blank)、[希腊字母](https://baike.baidu.com/item/%E5%B8%8C%E8%85%8A%E5%AD%97%E6%AF%8D" \t "https://baike.baidu.com/item/%E4%BF%A1%E6%81%AF%E4%BA%A4%E6%8D%A2%E7%94%A8%E6%B1%89%E5%AD%97%E7%BC%96%E7%A0%81%E5%AD%97%E7%AC%A6%E9%9B%86/_blank)、日文平假名及片假名字母、俄语西里尔字母在内的682个[全角字符](https://baike.baidu.com/item/%E5%85%A8%E8%A7%92%E5%AD%97%E7%AC%A6" \t "https://baike.baidu.com/item/%E4%BF%A1%E6%81%AF%E4%BA%A4%E6%8D%A2%E7%94%A8%E6%B1%89%E5%AD%97%E7%BC%96%E7%A0%81%E5%AD%97%E7%AC%A6%E9%9B%86/_blank)。GB 2312的出现，基本满足了汉字的计算机处理需要，它所收录的汉字已经覆盖中国大陆99.75%的使用频率。

字节数：GB 2312的程序通常采用[EUC](https://zh.wikipedia.org/wiki/EUC" \o "EUC)储存方法，以便兼容于[ASCII](https://zh.wikipedia.org/wiki/ASCII" \o "ASCII)。这种格式称为**EUC-CN**。[浏览器](https://zh.wikipedia.org/wiki/%E7%BD%91%E9%A1%B5%E6%B5%8F%E8%A7%88%E5%99%A8" \o "网页浏览器)编码表上的“GB2312”就是指这种表示法。每个汉字及符号以两个[字节](https://zh.wikipedia.org/wiki/%E5%AD%97%E8%8A%82" \o "字节)来表示。第一个字节称为“高位字节”，第二个字节称为“低位字节”。“高位字节”使用了0xA1–0xF7（把01–87区的区号加上0xA0），“低位字节”使用了0xA1–0xFE（把01–94加上0xA0）。 由于一级汉字从16区起始，汉字区的“高位字节”的范围是0xB0–0xF7，“低位字节”的范围是0xA1–0xFE，占用的码位是72\*94=6768。其中有5个空位是D7FA–D7FE。

兼容：有两种不同的GB 2312实现，在它们之间存在少量的差别，其中至少有一个是错误的。

|  |  |  |  |
| --- | --- | --- | --- |
| **字节序** | **GBK子集** | **GB2312.TXT** | **字符名称[[2]](https://zh.wikipedia.org/wiki/GB_2312" \l "cite_note-gb2312-80-3)** |
| A1A4 | U+00B7“·” MIDDLE DOT | U+30FB “・” KATAKANA MIDDLE DOT | [间隔点](https://zh.wikipedia.org/wiki/%E9%97%B4%E9%9A%94%E7%82%B9" \o "间隔点) |
| A1AA | U+2014“—” EM DASH | U+2015 “―” HORIZONTAL BAR | [破折号](https://zh.wikipedia.org/wiki/%E7%A0%B4%E6%8A%98%E5%8F%B7" \o "破折号) |

GBK子集与GBK/GB 18030兼容，GB2312.TXT则不兼容。后者基于ftp.unicode.org曾经提供的GB2312.TXT实现，于2011年由官方弃用，2016年9月时已无原文件踪迹。此外还有很多种厂商实现。 截至2015年 (2015-Missing required parameter 1=*month*!)[[update]](https://zh.wikipedia.org/w/index.php?title=GB_2312&action=edit)，微软.NET使用的是“GBK子集”实现。[ICU](https://zh.wikipedia.org/w/index.php?title=%E4%B8%87%E5%9B%BD%E7%A0%81%E5%9B%BD%E9%99%85%E5%8C%96%E7%BB%84%E4%BB%B6&action=edit&redlink=1)（英语：[International Components for Unicode](https://en.wikipedia.org/wiki/International_Components_for_Unicode" \o "en:International Components for Unicode)）、libiconv-1.14、php-5.6、ActivePerl-5.20、Java 1.7、Python 3.4都使用“GB2312.TXT”实现。Ruby 2.2兼容两者编码，但内部使用“GBK子集”实现。W3C的编码技术指南规定，应将GB2312字节流视为GBK编码，与GB18030一并使用同一解码器解码。

**2.BIG5:**

**来历：**大五码”（Big5）是由台湾[财团法人资讯工业策进会](https://zh.wikipedia.org/wiki/%E8%B3%87%E8%A8%8A%E5%B7%A5%E6%A5%AD%E7%AD%96%E9%80%B2%E6%9C%83" \o "资讯工业策进会)为[五大中文套装软体](https://zh.wikipedia.org/wiki/%E4%BA%94%E5%A4%A7%E4%B8%AD%E6%96%87%E5%A5%97%E8%A3%9D%E8%BB%9F%E9%AB%94" \o "五大中文套装软体)所设计的中文共通内码，在1983年12月完成公告，隔年3月，资讯工业策进会与[台湾](https://zh.wikipedia.org/wiki/%E8%87%BA%E7%81%A3" \o "台湾)13家厂商签定“16位元个人电脑套装软体合作开发（BIG-5）计画（[五大中文套装软体](https://zh.wikipedia.org/wiki/%E4%BA%94%E5%A4%A7%E4%B8%AD%E6%96%87%E5%A5%97%E8%A3%9D%E8%BB%9F%E9%AB%94" \o "五大中文套装软体)）”，因为此中文内码是为台湾自行制作开发之“五大中文套装软体”所设计的，所以就称为Big5中文内码。五大中文套装软体虽然并没有如预期的取代国外的套装软体，但随著采用Big5码的[国乔中文系统](https://zh.wikipedia.org/w/index.php?title=%E5%9C%8B%E5%96%AC%E4%B8%AD%E6%96%87%E7%B3%BB%E7%B5%B1&action=edit&redlink=1" \o "国乔中文系统（页面不存在）)及[倚天中文系统](https://zh.wikipedia.org/wiki/%E5%80%9A%E5%A4%A9%E4%B8%AD%E6%96%87%E7%B3%BB%E7%B5%B1" \o "倚天中文系统)先后在台湾市场获得成功，使得Big5码深远地影响正体中文电脑[内码](https://zh.wikipedia.org/wiki/%E5%85%A7%E7%A2%BC" \o "内码)，直至今日。“五大码”的英文名称“Big5”后来被人按英文字序译回中文，以致现在有“五大码”和“大五码”两个中文名称。 **产生原因：**Big5码的产生，是因为当时[个人电脑](https://zh.wikipedia.org/wiki/%E5%80%8B%E4%BA%BA%E9%9B%BB%E8%85%A6" \o "个人电脑)没有共通的内码，导致厂商推出的中文应用软体无法推广，并且与[IBM 5550](https://zh.wikipedia.org/wiki/IBM_5550" \o "IBM 5550)、[王安码](https://zh.wikipedia.org/w/index.php?title=%E7%8E%8B%E5%AE%89%E7%A2%BC&action=edit&redlink=1" \o "王安码（页面不存在）)等内码，彼此不能兼容；另一方面，台湾当时尚未推出中文编码标准。在这样的时空背景下，为了使台湾早日进入资讯时代，所采行的一个计画；同时，这个计画对于以台湾为核心的亚洲繁体汉字圈也产生了久远的影响。

Big5产生前，研发[中文电脑](https://zh.wikipedia.org/wiki/%E4%B8%AD%E6%96%87%E9%9B%BB%E8%85%A6" \o "中文电脑)的[朱邦复](https://zh.wikipedia.org/wiki/%E6%9C%B1%E9%82%A6%E5%BE%A9" \o "朱邦复)认为内码字集应该广纳所有的正异体字，以顾及如户政等应用上的需要，故在当时的内码会议中，建议希望采用他的五万多字的字库。工程师认为虽其技术可行，但是三个位元组（超过两个位元组）长度的内码却会造成英文萤幕画面映射成中文画面会发生文字无法对齐的问题，因为当时盛行之倚天中文系统画面系以两个位元组文字宽度映射成一个中文字图样，英文软体中只要以两个英文字宽度去显示一个中文字，画面就不会乱掉，造成中文系统业者偏爱二个位元组长度的内码；此外以仓颉输入码压缩成的内码不具排序等功能，因此未被采用。1983年有人诬指朱邦复为[共产党](https://zh.wikipedia.org/wiki/%E5%85%B1%E7%94%A2%E9%BB%A8" \o "共产党)，其研究成果更不可能获采用。在Big5码诞生后，大部分台湾的电脑软体都使用了Big5码，加上后来[倚天中文系统](https://zh.wikipedia.org/wiki/%E5%80%9A%E5%A4%A9%E4%B8%AD%E6%96%87%E7%B3%BB%E7%B5%B1" \o "倚天中文系统)的高度普及，使后来的[微软](https://zh.wikipedia.org/wiki/%E5%BE%AE%E8%BD%AF" \o "微软)[Windows 3.x](https://zh.wikipedia.org/wiki/Windows_3.x" \o "Windows 3.x)等亦予以采用。虽然后来台湾还有各种想要取代Big5码，像是倚天中文系统所推行的[倚天码](https://zh.wikipedia.org/wiki/%E5%80%9A%E5%A4%A9%E7%A2%BC" \o "倚天码)、台北市电脑公会所推动的[公会码](https://zh.wikipedia.org/w/index.php?title=%E5%85%AC%E6%9C%83%E7%A2%BC&action=edit&redlink=1" \o "公会码（页面不存在）)等，但是由于Big5字码已沿用多年，因此在习惯不易改变的情况下，始终无法成为主流字码。而台湾后来发展的[国家标准CNS 11643中文标准交换码](https://zh.wikipedia.org/wiki/CNS11643" \o "CNS11643)由于非一般的内码系统，是以交换使用为目的，受先天所限，必须使用至少三个位元组来表示一个汉字，所以普及率远远不及Big5码。在1990年代初期，当[中国大陆](https://zh.wikipedia.org/wiki/%E4%B8%AD%E5%9C%8B%E5%A4%A7%E9%99%B8" \o "中国大陆)的[电邮](https://zh.wikipedia.org/wiki/%E9%9B%BB%E9%83%B5" \o "电邮)和转码软体还未普遍之时，在[深圳](https://zh.wikipedia.org/wiki/%E6%B7%B1%E5%9C%B3" \o "深圳)的港商和台商公司亦曾经使用Big5系统，以方便与总部的文件交流、以及避免为大陆的办公室再写一套不同内码的系统。使用[简体中文](https://zh.wikipedia.org/wiki/%E7%AE%80%E4%BD%93%E4%B8%AD%E6%96%87" \o "简体中文)的社群，最常用的是[GB 2312](https://zh.wikipedia.org/wiki/GB_2312" \o "GB 2312)、[GBK](https://zh.wikipedia.org/wiki/GBK" \o "GBK)及其后续的[国标码](https://zh.wikipedia.org/wiki/%E5%9B%BD%E5%AE%B6%E6%A0%87%E5%87%86%E4%BB%A3%E7%A0%81" \o "国家标准代码)（[GB 18030](https://zh.wikipedia.org/wiki/GB_18030" \o "GB 18030)）。除了台湾外，其他使用繁体汉字的地区，如[香港](https://zh.wikipedia.org/wiki/%E9%A6%99%E6%B8%AF" \o "香港)（[香港增补字符集](https://zh.wikipedia.org/wiki/%E9%A6%99%E6%B8%AF%E5%A2%9E%E8%A3%9C%E5%AD%97%E7%AC%A6%E9%9B%86" \o "香港增补字符集)）、[澳门](https://zh.wikipedia.org/wiki/%E6%BE%B3%E9%96%80" \o "澳门)，及使用繁体汉字的海外华人，都曾普遍使用Big5码做为中文内码及交换码。

**字节数**：Big5码是一套[双位元组字符集](https://zh.wikipedia.org/wiki/%E5%8F%8C%E5%AD%97%E8%8A%82%E5%AD%97%E7%AC%A6%E9%9B%86" \o "双字节字符集)，使用了双八码储存方法，以两个字节来安放一个字。第一个字节称为“高位字节”，第二个字节称为“低位字节”。

“高位字节”使用了0x81-0xFE，“低位字节”使用了0x40-0x7E，及0xA1-0xFE。在Big5的分区中：

|  |  |
| --- | --- |
| 0x8140-0xA0FE | 保留给使用者自定义字元（[造字](https://zh.wikipedia.org/wiki/%E9%80%A0%E5%AD%97" \o "造字)区） |
| 0xA140-0xA3BF | [标点符号](https://zh.wikipedia.org/wiki/%E6%A8%99%E9%BB%9E%E7%AC%A6%E8%99%9F" \o "标点符号)、[希腊字母](https://zh.wikipedia.org/wiki/%E5%B8%8C%E8%85%8A%E5%AD%97%E6%AF%8D" \o "希腊字母)及特殊符号， 包括在0xA259-0xA261，安放了九个[计量用汉字](https://zh.wikipedia.org/wiki/%E8%A8%88%E9%87%8F%E7%94%A8%E6%BC%A2%E5%AD%97" \o "计量用汉字)：兙兛兞兝兡兣嗧瓩糎。 |
| 0xA3C0-0xA3FE | 保留。此区没有开放作造字区用。 |
| 0xA440-0xC67E | [常用汉字](https://zh.wikipedia.org/wiki/%E5%B8%B8%E7%94%A8%E6%BC%A2%E5%AD%97" \o "常用汉字)，先按[笔划](https://zh.wikipedia.org/wiki/%E7%AD%86%E5%8A%83" \o "笔划)再按[部首](https://zh.wikipedia.org/wiki/%E9%83%A8%E9%A6%96" \o "部首)排序。 |
| 0xC6A1-0xC8FE | 保留给使用者自定义字元（造字区） |
| 0xC940-0xF9D5 | [次常用汉字](https://zh.wikipedia.org/wiki/%E6%AC%A1%E5%B8%B8%E7%94%A8%E5%AD%97" \o "次常用字)，亦是先按笔划再按部首排序。 |
| 0xF9D6-0xFEFE | 保留给使用者自定义字元（造字区） |

值得留意的是，Big5重复收录了两个相同的字：“兀、兀”（0xA461[U+5140]及0xC94A[U+FA0C]）、“嗀、嗀”（0xDCD1[U+55C0]及0xDDFC[U+FA0D]）。此外“十”、“卅”也在符号区又重复了一次，在检索系统中常会造成查询不到字。

**描述：Big5**，又称为**大五码**或**五大码**，是使用[繁体中文](https://zh.wikipedia.org/wiki/%E7%B9%81%E4%BD%93%E4%B8%AD%E6%96%87" \o "繁体中文)（正体中文）社群中最常用的电脑[汉字](https://zh.wikipedia.org/wiki/%E6%B1%89%E5%AD%97" \o "汉字)[字符集](https://zh.wikipedia.org/wiki/%E5%AD%97%E7%AC%A6%E9%9B%86" \o "字符集)标准，共收录13,060个汉字[[1]](https://zh.wikipedia.org/wiki/%E5%A4%A7%E4%BA%94%E7%A2%BC" \l "cite_note-1)。

中文码分为[内码](https://zh.wikipedia.org/wiki/%E5%85%A7%E7%A2%BC" \o "内码)及[交换码](https://zh.wikipedia.org/wiki/%E4%BA%A4%E6%8F%9B%E7%A2%BC" \o "交换码)两类，Big5属中文内码，知名的中文交换码有[CCCII](https://zh.wikipedia.org/wiki/CCCII" \o "CCCII)、[CNS11643](https://zh.wikipedia.org/wiki/CNS11643" \o "CNS11643)。

Big5虽普及于[台湾](https://zh.wikipedia.org/wiki/%E5%8F%B0%E7%81%A3" \o "台湾)、[香港](https://zh.wikipedia.org/wiki/%E9%A6%99%E6%B8%AF" \o "香港)与[澳门](https://zh.wikipedia.org/wiki/%E6%BE%B3%E9%96%80" \o "澳门)等繁体中文通行区，但长期以来并非当地的国家/地区标准或官方标准，而只是**[业界标准](https://zh.wikipedia.org/wiki/De_facto" \o "De facto)**。[倚天中文系统](https://zh.wikipedia.org/wiki/%E5%80%9A%E5%A4%A9%E4%B8%AD%E6%96%87%E7%B3%BB%E7%B5%B1" \o "倚天中文系统)、[Windows](https://zh.wikipedia.org/wiki/Windows" \o "Windows)繁体中文版等主要系统的字符集都是以Big5为基准，但厂商又各自增加不同的造字与造字区，衍生成多种不同版本。

2003年，Big5被收录到CNS11643中文标准交换码的附录当中，取得了较正式的地位。这个最新版本被称为[Big5-2003](https://zh.wikipedia.org/w/index.php?title=Big5-2003&action=edit&redlink=1" \o "Big5-2003（页面不存在）)。

**兼容性：**因为低位元字元中包含了[程式语言](https://zh.wikipedia.org/wiki/%E7%A8%8B%E5%BC%8F%E8%AA%9E%E8%A8%80" \o "程式语言)、[shell](https://zh.wikipedia.org/wiki/Unix_shell" \o "Unix shell)、[script](https://zh.wikipedia.org/wiki/%E8%85%B3%E6%9C%AC%E8%AA%9E%E8%A8%80" \o "脚本语言)中，字串或命令常会用到的特殊字元，例如0x5C“\”、0x7C“|”等。“\”在许多用途的字串中是当作转义符号又称为跳脱字元，例如\n（换行）、\r（归位)、\t（tab）、\\（\本身符号）、\"（引号）等等。而“|”在[UNIX](https://zh.wikipedia.org/wiki/UNIX" \o "UNIX)作业系统中大多当作命令管线的使用，如"ls -la | more"等等。如果在字串中有这些特殊的转义字元，会被[程式](https://zh.wikipedia.org/wiki/%E7%A8%8B%E5%BC%8F" \o "程式)或[直译器](https://zh.wikipedia.org/wiki/%E7%9B%B4%E8%AD%AF%E5%99%A8" \o "直译器)解释为特殊用途。但是因为是中文的原因，故无法正确解释为上面所述的行为，因此程式可能会忽略此转义符号或是中断执行。若此，就违反了使用者本来要当成[中文字元](https://zh.wikipedia.org/w/index.php?title=%E4%B8%AD%E6%96%87%E5%AD%97%E5%85%83&action=edit&redlink=1" \o "中文字元（页面不存在）)一部份使用的本意。

**UTF-16:**

**描述：**Unicode的编码空间从U+0000到U+10FFFF，共有1,112,064个码位（code point）可用来映射字符. Unicode的编码空间可以划分为17个平面（plane），每个平面包含216（65,536）个码位。17个平面的码位可表示为从U+xx0000到U+xxFFFF，其中xx表示十六进制值从0016到1016，共计17个平面。第一个平面称为**基本多语言平面**（Basic Multilingual Plane, **BMP**），或称第零平面（Plane 0）。其他平面称为**辅助平面**（Supplementary Planes）。基本多语言平面内，从U+D800到U+DFFF之间的码位区段是永久保留不映射到Unicode字符。UTF-16就利用保留下来的0xD800-0xDFFF区段的码位来对辅助平面的字符的码位进行编码。

### 从U+0000至U+D7FF以及从U+E000至U+FFFF的码位第一个Unicode平面（码位从U+0000至U+FFFF）包含了最常用的字符。该平面被称为基本多语言平面，缩写为*BMP*（Basic Multilingual Plane, BMP）。UTF-16与[UCS-2](https://zh.wikipedia.org/wiki/UCS-2" \o "UCS-2)编码这个范围内的码位为16比特长的单个码元，数值等价于对应的码位. BMP中的这些码位是仅有的可以在UCS-2中表示的码位。

### 从U+10000到U+10FFFF的码位

辅助平面（Supplementary Planes）中的码位，在UTF-16中被编码为**一对**16比特长的码元（即32bit,4Bytes），称作*代理对*（surrogate pair），具体方法是：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| UTF-16解码 | | | | |
| lead \ trail | **DC00** | **DC01** | **…** | **DFFF** |
| **D800** | 10000 | 10001 | … | 103FF |
| **D801** | 10400 | 10401 | … | 107FF |
| **⋮** | ⋮ | ⋮ | ⋱ | ⋮ |
| **DBFF** | 10FC00 | 10FC01 | … | 10FFFF |

* 码位减去0x10000,得到的值的范围为20比特长的0..0xFFFFF.
* 高位的10比特的值（值的范围为0..0x3FF）被加上0xD800得到第一个码元或称作高位代理（high surrogate），值的范围是0xD800..0xDBFF.由于高位代理比低位代理的值要小，所以为了避免混淆使用，Unicode标准现在称高位代理为**前导代理**（lead surrogates）。
* 低位的10比特的值（值的范围也是0..0x3FF）被加上0xDC00得到第二个码元或称作低位代理（low surrogate），现在值的范围是0xDC00..0xDFFF.由于低位代理比高位代理的值要大，所以为了避免混淆使用，Unicode标准现在称低位代理为**后尾代理**（trail surrogates）。

**Unicode：**

**描述：Unicode**（[中文](https://zh.wikipedia.org/wiki/%E4%B8%AD%E6%96%87" \o "中文)：万国码、国际码、统一码、单一码）是[电脑科学](https://zh.wikipedia.org/wiki/%E9%9B%BB%E8%85%A6%E7%A7%91%E5%AD%B8" \o "电脑科学)领域里的一项业界标准。它对世界上大部分的[文字系统](https://zh.wikipedia.org/wiki/%E6%96%87%E5%AD%97%E7%B3%BB%E7%B5%B1" \o "文字系统)进行了整理、编码，使得电脑可以用更为简单的方式来呈现和处理文字。

Unicode伴随著[通用字符集](https://zh.wikipedia.org/wiki/%E9%80%9A%E7%94%A8%E5%AD%97%E7%AC%A6%E9%9B%86" \o "通用字符集)的标准而发展，同时也以书本的形式[[1]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-1)对外发表。Unicode至今仍在不断增修，每个新版本都加入更多新的字符。目前最新的版本为2017年6月20日公布的10.0.0[[2]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-Unicode10.0-2)，已经收录超过十万个[字符](https://zh.wikipedia.org/wiki/%E5%AD%97%E7%AC%A6" \o "字符)（第十万个字符在2005年获采纳）。Unicode涵盖的资料除了视觉上的字形、编码方法、标准的[字符编码](https://zh.wikipedia.org/wiki/%E5%AD%97%E7%AC%A6%E7%BC%96%E7%A0%81" \o "字符编码)外，还包含了字符特性，如大小写字母。

Unicode发展由非营利机构[统一码联盟](https://zh.wikipedia.org/wiki/%E7%B5%B1%E4%B8%80%E7%A2%BC%E8%81%AF%E7%9B%9F" \o "统一码联盟)负责，该机构致力于让Unicode方案取代既有的字符编码方案。因为既有的方案往往空间非常有限，亦不适用于[多语](https://zh.wikipedia.org/wiki/%E5%A4%9A%E8%AA%9E" \o "多语)环境。

Unicode备受认可，并广泛地应用于电脑软体的[国际化与本地化](https://zh.wikipedia.org/wiki/%E5%9C%8B%E9%9A%9B%E5%8C%96%E8%88%87%E6%9C%AC%E5%9C%B0%E5%8C%96" \o "国际化与本地化)过程。有很多新科技，如[可扩展置标语言](https://zh.wikipedia.org/wiki/%E5%8F%AF%E6%89%A9%E5%B1%95%E7%BD%AE%E6%A0%87%E8%AF%AD%E8%A8%80" \o "可扩展置标语言)(Extensible Markup Language，简称：XML)、[Java程式语言](https://zh.wikipedia.org/wiki/Java" \o "Java)以及现代的[作业系统](https://zh.wikipedia.org/wiki/%E4%BD%9C%E6%A5%AD%E7%B3%BB%E7%B5%B1" \o "作业系统)，都采用Unicode编码。

**产生原因：**Unicode是为了解决传统的[字元编码](https://zh.wikipedia.org/wiki/%E5%AD%97%E5%85%83%E7%B7%A8%E7%A2%BC" \o "字元编码)方案的局限而产生的，例如[ISO 8859-1](https://zh.wikipedia.org/wiki/ISO/IEC_8859" \o "ISO/IEC 8859)所定义的字元虽然在不同的国家中广泛地使用，可是在不同国家间却经常出现不相容的情况。很多传统的编码方式都有一个共同的问题，即容许电脑处理双语环境（通常使用[拉丁字母](https://zh.wikipedia.org/wiki/%E6%8B%89%E4%B8%81%E5%AD%97%E6%AF%8D" \o "拉丁字母)以及其本地语言），但却无法同时支援多语言环境（指可同时处理多种语言混合的情况）。

Unicode编码包含了不同写法的字，如“ɑ／a”、“強／强”、“戶／户／戸”。然而在[汉字](https://zh.wikipedia.org/wiki/%E6%B1%89%E5%AD%97" \o "汉字)方面引起了一字多形的认定争议（详见[中日韩统一表意文字](https://zh.wikipedia.org/wiki/%E4%B8%AD%E6%97%A5%E9%9F%93%E7%B5%B1%E4%B8%80%E8%A1%A8%E6%84%8F%E6%96%87%E5%AD%97" \o "中日韩统一表意文字)主题）。

在文字处理方面，统一码为每一个字符而非字形定义唯一的代码（即一个整数）。换句话说，统一码以一种抽象的方式（即数字）来处理字符，并将视觉上的演绎工作（例如字体大小、外观形状、字体形态、文体等）留给其他软件来处理，例如网页浏览器或是文字处理器。

目前，几乎所有电脑系统都支持基本拉丁字母，并各自支持不同的其他编码方式。Unicode为了和它们相互兼容，其首256字元保留给ISO 8859-1所定义的字元，使既有的西欧语系文字的转换不需特别考量；并且把大量相同的字元重复编到不同的字元码中去，使得旧有纷杂的编码方式得以和Unicode编码间互相直接转换，而不会遗失任何资讯。举例来说，[全形](https://zh.wikipedia.org/wiki/%E5%85%A8%E5%BD%A2" \o "全形)格式区段包含了主要的拉丁字母的全形格式，在中文、日文、以及韩文字形当中，这些字元以全形的方式来呈现，而不以常见的半形形式显示，这对竖排文字和等宽排列文字有重要作用。

在表示一个Unicode的字元时，通常会用“U+”然后紧接着一组十六进位的数字来表示这一个字元。在[基本多文种平面](https://zh.wikipedia.org/wiki/%E5%9F%BA%E6%9C%AC%E5%A4%9A%E6%96%87%E7%A8%AE%E5%B9%B3%E9%9D%A2" \o "基本多文种平面)（英文：Basic Multilingual Plane，简写BMP。又称为“零号平面”、plane 0）里的所有字元，要用四个数字（即两个char,16bit ,例如U+4AE0，共支持六万多个字符）；在零号平面以外的字元则需要使用五个或六个数字。旧版的Unicode标准使用相近的标记方法，但却有些微小差异：在Unicode 3.0里使用“U-”然后紧接着八个数字，而“U+”则必须随后紧接着四个数字。

### 标准

位于美国加州的Unicode组织允许任何愿意支付会费的公司和个人加入，其成员包含了主要的电脑软硬体厂商，例如[奥多比系统](https://zh.wikipedia.org/wiki/Adobe_Systems" \o "Adobe Systems)、[苹果公司](https://zh.wikipedia.org/wiki/%E8%98%8B%E6%9E%9C%E5%85%AC%E5%8F%B8" \o "苹果公司)、[惠普](https://zh.wikipedia.org/wiki/HP" \o "HP)、[IBM](https://zh.wikipedia.org/wiki/IBM" \o "IBM)、[微软](https://zh.wikipedia.org/wiki/%E5%BE%AE%E8%BB%9F" \o "微软)、[施乐](https://zh.wikipedia.org/wiki/%E6%96%BD%E4%B9%90" \o "施乐)等。

20世纪80年代末，组成Unicode组织的商业机构，和国际合作的[国际标准化组织](https://zh.wikipedia.org/wiki/%E5%9C%8B%E9%9A%9B%E6%A8%99%E6%BA%96%E5%8C%96%E7%B5%84%E7%B9%94" \o "国际标准化组织)因为电脑普及和资讯国际化的前提下，分别各自成立了Unicode组织[[3]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-3)和ISO-10646工作小组。他们不久便发现对方的存在，大家为著相同的目的而工作。1991年，Unicode Consortium与ISO/IEC JTC1/SC2同意保持Unicode码表与ISO 10646标准保持兼容并密切协调各自标准进一步的扩展。虽然实际上两者的字集编码相同，但实质上两者确实为两个不同的标准。Unicode 1.1对应于ISO 10646-1:1993，Unicode 3.0对应于ISO 10646-1:2000，Unicode 3.2对应于ISO 10646-2:2001，Unicode 4.0对应于ISO 10646:2003，Unicode 5.0对应于ISO 10646:2003及附录1–3。

Unicode自版本2.0开始保持了向后兼容，即新的版本仅仅增加字符，原有字符不会被删除或更名。

统一码联盟在1991年首次发布了**The Unicode Standard**。Unicode的开发结合了[国际标准化组织](https://zh.wikipedia.org/wiki/%E5%9C%8B%E9%9A%9B%E6%A8%99%E6%BA%96%E5%8C%96%E7%B5%84%E7%B9%94" \o "国际标准化组织)所制定的[ISO/IEC 10646](https://zh.wikipedia.org/wiki/ISO/IEC_10646" \o "ISO/IEC 10646)，即[通用字元集](https://zh.wikipedia.org/w/index.php?title=%E9%80%9A%E7%94%A8%E5%AD%97%E5%85%83%E9%9B%86&action=edit&redlink=1" \o "通用字元集（页面不存在）)。Unicode与ISO/IEC 10646在编码的运作原理相同，但**The Unicode Standard**包含了更详尽的实现资讯、涵盖了更细节的主题，诸如位元编码（bitwise encoding）、校对以及呈现等。**The Unicode Standard**也列举了诸多的字元特性，包含了那些必须支援两种阅读方向的文字（由左至右或由右至左的文字阅读方向，例如阿拉伯文是由右至左）。Unicode与ISO/IEC 10646这两个标准在术语上的使用有些微的不同。

在2005年，Unicode的第十万个字元被引入成为标准之一，该字元被用于[马拉雅拉姆语](https://zh.wikipedia.org/wiki/%E9%A6%AC%E6%8B%89%E9%9B%85%E6%8B%89%E5%A7%86%E8%AA%9E" \o "马拉雅拉姆语)。

### 历史

截至目前的Unicode各版本及其发布时间如下：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Unicode版本 | | | | | | |
| **版本** | **发布日期** | **书籍** | **对应[ISO/IEC 10646](https://zh.wikipedia.org/wiki/%E9%80%9A%E7%94%A8%E5%AD%97%E7%AC%A6%E9%9B%86" \o "通用字符集)版本** | **[字母](https://zh.wikipedia.org/w/index.php?title=%E5%AD%97%E6%AF%8D_(Unicode)&action=edit&redlink=1)（英语：[Script (Unicode)](https://en.wikipedia.org/wiki/Script_(Unicode)" \o "en:Script (Unicode))）** | **字元** | |
| **总计[[a]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-4)** | **已知的扩增** |
| 1.0.0 | 1991年10月 | [ISBN 978-0-201-56788-5](https://zh.wikipedia.org/wiki/Special:%E7%BD%91%E7%BB%9C%E4%B9%A6%E6%BA%90/9780201567885) (Vol.1) |  | 24 | 7,161 | 最初包含的文字有：[阿拉伯字母](https://zh.wikipedia.org/wiki/%E9%98%BF%E6%8B%89%E4%BC%AF%E5%AD%97%E6%AF%8D" \o "阿拉伯字母)、[亚美尼亚字母](https://zh.wikipedia.org/wiki/%E4%BA%9E%E7%BE%8E%E5%B0%BC%E4%BA%9E%E5%AD%97%E6%AF%8D" \o "亚美尼亚字母)、[孟加拉文](https://zh.wikipedia.org/wiki/%E5%AD%9F%E5%8A%A0%E6%8B%89%E6%96%87" \o "孟加拉文)、[注音符号](https://zh.wikipedia.org/wiki/%E6%B3%A8%E9%9F%B3%E7%AC%A6%E8%99%9F" \o "注音符号)、[西里尔字母](https://zh.wikipedia.org/wiki/%E8%A5%BF%E9%87%8C%E7%88%BE%E5%AD%97%E6%AF%8D" \o "西里尔字母)、[天城文](https://zh.wikipedia.org/wiki/%E5%A4%A9%E5%9F%8E%E6%96%87" \o "天城文)、[乔治亚字母](https://zh.wikipedia.org/wiki/%E5%96%AC%E6%B2%BB%E4%BA%9E%E5%AD%97%E6%AF%8D" \o "乔治亚字母)、[希腊字母](https://zh.wikipedia.org/wiki/%E5%B8%8C%E8%87%98%E5%AD%97%E6%AF%8D" \o "希腊字母)、[古吉拉特文](https://zh.wikipedia.org/wiki/%E5%8F%A4%E5%90%89%E6%8B%89%E7%89%B9%E6%96%87" \o "古吉拉特文)、[古木基文](https://zh.wikipedia.org/wiki/%E5%8F%A4%E6%9C%A8%E5%9F%BA%E6%96%87" \o "古木基文)、[谚文](https://zh.wikipedia.org/wiki/%E8%AB%BA%E6%96%87" \o "谚文)、[希伯来字母](https://zh.wikipedia.org/wiki/%E5%B8%8C%E4%BC%AF%E4%BE%86%E5%AD%97%E6%AF%8D" \o "希伯来字母)、[平假名](https://zh.wikipedia.org/wiki/%E5%B9%B3%E5%81%87%E5%90%8D" \o "平假名)、[卡纳达文](https://zh.wikipedia.org/wiki/%E5%8D%A1%E7%B4%8D%E9%81%94%E6%96%87" \o "卡纳达文)、[片假名](https://zh.wikipedia.org/wiki/%E7%89%87%E5%81%87%E5%90%8D" \o "片假名)、[寮文字](https://zh.wikipedia.org/wiki/%E5%AF%AE%E6%96%87%E5%AD%97" \o "寮文字)、[拉丁字母](https://zh.wikipedia.org/wiki/%E6%8B%89%E4%B8%81%E5%AD%97%E6%AF%8D" \o "拉丁字母)、[马拉雅拉姆文](https://zh.wikipedia.org/wiki/%E9%A6%AC%E6%8B%89%E9%9B%85%E6%8B%89%E5%A7%86%E6%96%87" \o "马拉雅拉姆文)、[奥里亚文](https://zh.wikipedia.org/wiki/%E5%A5%A7%E9%87%8C%E4%BA%9E%E6%96%87" \o "奥里亚文)、[泰米尔文](https://zh.wikipedia.org/wiki/%E6%B3%B0%E7%B1%B3%E7%88%BE%E6%96%87" \o "泰米尔文)、[泰卢固文](https://zh.wikipedia.org/wiki/%E6%B3%B0%E5%8D%A2%E5%9B%BA%E6%96%87" \o "泰卢固文)、[泰文字](https://zh.wikipedia.org/wiki/%E6%B3%B0%E6%96%87%E5%AD%97" \o "泰文字)与[藏文](https://zh.wikipedia.org/wiki/%E8%97%8F%E6%96%87" \o "藏文)[[4]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-5)。 |
| 1.0.1 | 1992年6月 | [ISBN 978-0-201-60845-8](https://zh.wikipedia.org/wiki/Special:%E7%BD%91%E7%BB%9C%E4%B9%A6%E6%BA%90/9780201608458) (Vol.2) |  | 25 | 28,359 | 定义[中日韩统一表意文字](https://zh.wikipedia.org/wiki/%E4%B8%AD%E6%97%A5%E9%9F%93%E7%B5%B1%E4%B8%80%E8%A1%A8%E6%84%8F%E6%96%87%E5%AD%97" \o "中日韩统一表意文字)最初的20,902个字[[5]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-6)。 |
| 1.1 | 1993年6月 |  | ISO/IEC 10646-1:1993 | 24 | 34,233 | 于原有2,350个[谚文](https://zh.wikipedia.org/wiki/%E8%AB%BA%E6%96%87" \o "谚文)字母的基础上新增4,306个谚文字母。移除[藏文](https://zh.wikipedia.org/wiki/%E8%97%8F%E6%96%87" \o "藏文)[[6]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-7)。 |
| 2.0 | 1996年7月 | [ISBN 978-0-201-48345-1](https://zh.wikipedia.org/wiki/Special:%E7%BD%91%E7%BB%9C%E4%B9%A6%E6%BA%90/9780201483451) | ISO/IEC 10646-1:1993与其第5－7修订版 | 25 | 38,950 | 移除原有的[谚文](https://zh.wikipedia.org/wiki/%E8%AB%BA%E6%96%87" \o "谚文)字母设置，于新的编码范围更换成11,172个新的谚文字母。[藏文](https://zh.wikipedia.org/wiki/%E8%97%8F%E6%96%87" \o "藏文)重新加入，但编码位置更换。代理字符机制建立，并将第15与第16平面分配给私人使用区[[7]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-8)。 |
| 2.1 | 1998年5月 |  | ISO/IEC 10646-1:1993与其第5－7修订版，以及第18修订版中新增的2个字元 | 25 | 38,952 | 新增[欧元符号](https://zh.wikipedia.org/wiki/%E6%AC%A7%E5%85%83%E7%AC%A6%E5%8F%B7" \o "欧元符号)与[对象替换字符](https://zh.wikipedia.org/w/index.php?title=%E5%AF%B9%E8%B1%A1%E6%9B%BF%E6%8D%A2%E5%AD%97%E7%AC%A6&action=edit&redlink=1" \o "对象替换字符（页面不存在）)[[8]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-9)。 |
| 3.0 | 1999年9月 | [ISBN 978-0-201-61633-0](https://zh.wikipedia.org/wiki/Special:%E7%BD%91%E7%BB%9C%E4%B9%A6%E6%BA%90/9780201616330) | ISO/IEC 10646-1:2000 | 38 | 49,259 | 新增[切罗基文](https://zh.wikipedia.org/wiki/%E5%88%87%E7%BD%97%E5%9F%BA%E6%96%87" \o "切罗基文)，[埃塞俄比亚语](https://zh.wikipedia.org/wiki/%E5%9F%83%E5%A1%9E%E4%BF%84%E6%AF%94%E4%BA%9E%E8%AA%9E" \o "埃塞俄比亚语)，[高棉语](https://zh.wikipedia.org/wiki/%E9%AB%98%E6%A3%89%E8%AF%AD" \o "高棉语)，[蒙古语](https://zh.wikipedia.org/wiki/%E8%92%99%E5%8F%A4%E8%AF%AD" \o "蒙古语)，[缅甸语](https://zh.wikipedia.org/wiki/%E7%BC%85%E7%94%B8%E8%AF%AD" \o "缅甸语)，[欧甘字母](https://zh.wikipedia.org/wiki/%E6%AD%90%E7%94%98%E5%AD%97%E6%AF%8D" \o "欧甘字母)，[卢恩字母](https://zh.wikipedia.org/wiki/%E5%8D%A2%E6%81%A9%E5%AD%97%E6%AF%8D" \o "卢恩字母)，[僧伽罗语](https://zh.wikipedia.org/wiki/%E5%83%A7%E4%BC%BD%E7%BD%97%E8%AF%AD" \o "僧伽罗语)，[叙利亚语](https://zh.wikipedia.org/wiki/%E6%95%98%E5%88%A9%E4%BA%9E%E8%AA%9E" \o "叙利亚语)，[它拿字母](https://zh.wikipedia.org/wiki/%E5%AE%83%E6%8B%BF%E5%AD%97%E6%AF%8D" \o "它拿字母)，[加拿大原住民音节文字](https://zh.wikipedia.org/wiki/%E5%8A%A0%E6%8B%BF%E5%A4%A7%E5%8E%9F%E4%BD%8F%E6%B0%91%E9%9F%B3%E7%AF%80%E6%96%87%E5%AD%97" \o "加拿大原住民音节文字)和[彝文](https://zh.wikipedia.org/wiki/%E5%BD%9D%E6%96%87" \o "彝文)，以及部分[盲文](https://zh.wikipedia.org/wiki/%E7%9B%B2%E6%96%87" \o "盲文)图案。[[9]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-10) |
| 3.1 | 2001年3月 |  | ISO/IEC 10646-1:2000  ISO/IEC 10646-2:2001 | 41 | 94,205 | 新增[犹他字母](https://zh.wikipedia.org/w/index.php?title=%E7%8A%B9%E4%BB%96%E5%AD%97%E6%AF%8D&action=edit&redlink=1" \o "犹他字母（页面不存在）)、[哥特字母](https://zh.wikipedia.org/wiki/%E5%93%A5%E7%89%B9%E5%AD%97%E6%AF%8D" \o "哥特字母)、[古意大利字母](https://zh.wikipedia.org/wiki/%E5%8F%A4%E6%84%8F%E5%A4%A7%E5%88%A9%E5%AD%97%E6%AF%8D" \o "古意大利字母)、[音乐符号](https://zh.wikipedia.org/wiki/%E9%9F%B3%E6%A8%82%E7%AC%A6%E8%99%9F" \o "音乐符号)和[拜占庭音乐符号](https://zh.wikipedia.org/wiki/%E6%8B%9C%E5%8D%A0%E5%BA%AD%E9%9F%B3%E4%B9%90%E7%AC%A6%E5%8F%B7" \o "拜占庭音乐符号)，追加了42711个[中日韩统一表意文字](https://zh.wikipedia.org/wiki/%E4%B8%AD%E6%97%A5%E9%9F%A9%E7%BB%9F%E4%B8%80%E8%A1%A8%E6%84%8F%E6%96%87%E5%AD%97" \o "中日韩统一表意文字)。[[10]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-11) |
| 3.2 | 2002年3月 |  | ISO/IEC 10646-1:2000与其第1修订版  ISO/IEC 10646-2:2001 | 45 | 95,221 | 新增[菲律宾](https://zh.wikipedia.org/wiki/%E8%8F%B2%E5%BE%8B%E5%AE%BE" \o "菲律宾)文字[布锡文](https://zh.wikipedia.org/wiki/%E5%B8%83%E9%94%A1%E6%96%87" \o "布锡文)、[哈努诺文](https://zh.wikipedia.org/w/index.php?title=%E5%93%88%E5%8A%AA%E8%AF%BA%E6%96%87&action=edit&redlink=1" \o "哈努诺文（页面不存在）)、[他加禄语](https://zh.wikipedia.org/wiki/%E4%BB%96%E5%8A%A0%E7%A6%84%E8%AF%AD" \o "他加禄语)、[塔格巴奴亚文](https://zh.wikipedia.org/w/index.php?title=%E5%A1%94%E6%A0%BC%E5%B7%B4%E5%A5%B4%E4%BA%9A%E6%96%87&action=edit&redlink=1" \o "塔格巴奴亚文（页面不存在）)。[[11]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-12) |
| 4.0 | 2003年4月 | [ISBN 978-0-321-18578-5](https://zh.wikipedia.org/wiki/Special:%E7%BD%91%E7%BB%9C%E4%B9%A6%E6%BA%90/9780321185785) | ISO/IEC 10646:2003 | 52 | 96,447 | 新增[塞浦路斯音节文字](https://zh.wikipedia.org/wiki/%E5%A1%9E%E6%B5%A6%E8%B7%AF%E6%96%AF%E9%9F%B3%E7%AF%80%E6%96%87%E5%AD%97" \o "塞浦路斯音节文字)，[林布字母](https://zh.wikipedia.org/w/index.php?title=%E6%9E%97%E5%B8%83%E5%AD%97%E6%AF%8D&action=edit&redlink=1" \o "林布字母（页面不存在）)，[线形文字B](https://zh.wikipedia.org/wiki/%E7%BA%BF%E5%BD%A2%E6%96%87%E5%AD%97B" \o "线形文字B)，[奥斯曼亚字母](https://zh.wikipedia.org/wiki/%E5%A5%A7%E6%96%AF%E6%9B%BC%E4%BA%9E%E5%AD%97%E6%AF%8D" \o "奥斯曼亚字母)，[萧伯纳字母](https://zh.wikipedia.org/wiki/%E8%95%AD%E4%BC%AF%E7%B4%8D%E5%AD%97%E6%AF%8D" \o "萧伯纳字母)，[德宏傣文](https://zh.wikipedia.org/wiki/%E5%BE%B7%E5%AE%8F%E5%82%A3%E6%96%87" \o "德宏傣文)，[乌加里特字母](https://zh.wikipedia.org/wiki/%E4%B9%8C%E5%8A%A0%E9%87%8C%E7%89%B9%E5%AD%97%E6%AF%8D" \o "乌加里特字母)以及[六十四卦](https://zh.wikipedia.org/wiki/%E5%85%AD%E5%8D%81%E5%9B%9B%E5%8D%A6" \o "六十四卦)。[[12]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-13) |
| 4.1 | 2005年3月 |  | ISO/IEC 10646:2003与其第1修订版 | 59 | 97,720 | 新增[布吉文](https://zh.wikipedia.org/w/index.php?title=%E5%B8%83%E5%90%89%E6%96%87&action=edit&redlink=1" \o "布吉文（页面不存在）)，[格拉哥里字母](https://zh.wikipedia.org/wiki/%E6%A0%BC%E6%8B%89%E5%93%A5%E9%87%8C%E5%AD%97%E6%AF%8D" \o "格拉哥里字母)，[佉卢文](https://zh.wikipedia.org/wiki/%E4%BD%89%E5%8D%A2%E6%96%87" \o "佉卢文)，[西双版纳傣文](https://zh.wikipedia.org/wiki/%E8%A5%BF%E5%8F%8C%E7%89%88%E7%BA%B3%E5%82%A3%E6%96%87" \o "西双版纳傣文)，[古波斯语](https://zh.wikipedia.org/wiki/%E5%8F%A4%E6%B3%A2%E6%96%AF%E8%AA%9E" \o "古波斯语)，[锡尔赫特文](https://zh.wikipedia.org/w/index.php?title=%E9%94%A1%E5%B0%94%E8%B5%AB%E7%89%B9%E6%96%87&action=edit&redlink=1" \o "锡尔赫特文（页面不存在）)和[提非纳文](https://zh.wikipedia.org/wiki/%E6%8F%90%E9%9D%9E%E7%B4%8D%E6%96%87" \o "提非纳文) 。[科普特字母](https://zh.wikipedia.org/wiki/%E7%A7%91%E6%99%AE%E7%89%B9%E5%AD%97%E6%AF%8D" \o "科普特字母)从[希腊语](https://zh.wikipedia.org/wiki/%E5%B8%8C%E8%85%8A%E8%AF%AD" \o "希腊语)区块中分离了出来。新增了[古希腊音乐符号](https://zh.wikipedia.org/w/index.php?title=%E5%8F%A4%E5%B8%8C%E8%85%8A%E9%9F%B3%E4%B9%90%E7%AC%A6%E5%8F%B7&action=edit&redlink=1" \o "古希腊音乐符号（页面不存在）)。[[13]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-14) |
| 5.0 | 2006年7月 | [ISBN 978-0-321-48091-0](https://zh.wikipedia.org/wiki/Special:%E7%BD%91%E7%BB%9C%E4%B9%A6%E6%BA%90/9780321480910) | ISO/IEC 10646:2003与其第1、2修订版，以及第3修订版中新增的4个字元 | 64 | 99,089 | 新增[巴厘语](https://zh.wikipedia.org/wiki/%E5%B7%B4%E5%8E%98%E8%AF%AD" \o "巴厘语)，[楔形文字](https://zh.wikipedia.org/wiki/%E6%A5%94%E5%BD%A2%E6%96%87%E5%AD%97" \o "楔形文字)，[西非书面文](https://zh.wikipedia.org/w/index.php?title=%E8%A5%BF%E9%9D%9E%E4%B9%A6%E9%9D%A2%E6%96%87&action=edit&redlink=1" \o "西非书面文（页面不存在）)，[八思巴文](https://zh.wikipedia.org/wiki/%E5%85%AB%E6%80%9D%E5%B7%B4%E6%96%87" \o "八思巴文)和[腓尼基字母](https://zh.wikipedia.org/wiki/%E8%85%93%E5%B0%BC%E5%9F%BA%E5%AD%97%E6%AF%8D" \o "腓尼基字母)。[[14]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-15) |
| 5.1 | 2008年4月 |  | ISO/IEC 10646:2003与其第1－4修订版 | 75 | 100,713 | 新增[卡利亚语](https://zh.wikipedia.org/w/index.php?title=%E5%8D%A1%E5%88%A9%E4%BA%9A%E8%AF%AD&action=edit&redlink=1" \o "卡利亚语（页面不存在）)，[占婆字母](https://zh.wikipedia.org/wiki/%E5%8D%A0%E5%A9%86%E5%AD%97%E6%AF%8D" \o "占婆字母)，[克耶黎语](https://zh.wikipedia.org/w/index.php?title=%E5%85%8B%E8%80%B6%E9%BB%8E%E8%AF%AD&action=edit&redlink=1" \o "克耶黎语（页面不存在）)，[雷布查语](https://zh.wikipedia.org/w/index.php?title=%E9%9B%B7%E5%B8%83%E6%9F%A5%E8%AF%AD&action=edit&redlink=1" \o "雷布查语（页面不存在）)，[利西亚语](https://zh.wikipedia.org/w/index.php?title=%E5%88%A9%E8%A5%BF%E4%BA%9A%E8%AF%AD&action=edit&redlink=1" \o "利西亚语（页面不存在）)，[吕底亚语](https://zh.wikipedia.org/wiki/%E5%91%82%E5%BA%95%E4%BA%9E%E8%AA%9E" \o "吕底亚语)，[桑塔利文](https://zh.wikipedia.org/w/index.php?title=%E6%A1%91%E5%A1%94%E5%88%A9%E6%96%87&action=edit&redlink=1" \o "桑塔利文（页面不存在）)，[拉让文](https://zh.wikipedia.org/w/index.php?title=%E6%8B%89%E8%AE%A9%E6%96%87&action=edit&redlink=1" \o "拉让文（页面不存在）)，[索拉什特拉文](https://zh.wikipedia.org/w/index.php?title=%E7%B4%A2%E6%8B%89%E4%BB%80%E7%89%B9%E6%8B%89%E6%96%87&action=edit&redlink=1" \o "索拉什特拉文（页面不存在）)，[巽他语](https://zh.wikipedia.org/wiki/%E5%B7%BD%E4%BB%96%E8%AA%9E" \o "巽他语)和[瓦伊语](https://zh.wikipedia.org/wiki/%E7%93%A6%E4%BC%8A%E8%AA%9E" \o "瓦伊语)。同时增加了[斐斯托斯圆盘](https://zh.wikipedia.org/wiki/%E6%96%90%E6%96%AF%E6%89%98%E6%96%AF%E5%9C%93%E7%9B%A4" \o "斐斯托斯圆盘)，[麻将](https://zh.wikipedia.org/wiki/%E9%BA%BB%E5%B0%86" \o "麻将)和[多米诺骨牌](https://zh.wikipedia.org/wiki/%E5%A4%9A%E7%B1%B3%E8%AF%BA%E9%AA%A8%E7%89%8C" \o "多米诺骨牌)上的符号。对[缅甸语](https://zh.wikipedia.org/wiki/%E7%BC%85%E7%94%B8%E8%AF%AD" \o "缅甸语)做了重要的补充，追加了手抄缩写的额外字母，追加了[大写ẞ](https://zh.wikipedia.org/wiki/%E5%A4%A7%E5%AF%AB%E1%BA%9E" \o "大写ẞ)。[[15]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-16) |
| 5.2 | 2009年10月 |  | ISO/IEC 10646:2003与其第1－6修订版 | 90 | 107,361 | 新增[阿维斯陀语](https://zh.wikipedia.org/wiki/%E9%98%BF%E7%BB%B4%E6%96%AF%E9%99%80%E8%AF%AD" \o "阿维斯陀语)，[巴姆文](https://zh.wikipedia.org/w/index.php?title=%E5%B7%B4%E5%A7%86%E6%96%87&action=edit&redlink=1" \o "巴姆文（页面不存在）)，[埃及象形文字](https://zh.wikipedia.org/wiki/%E5%9F%83%E5%8F%8A%E8%B1%A1%E5%BD%A2%E6%96%87%E5%AD%97" \o "埃及象形文字) （[加汀纳符号表](https://zh.wikipedia.org/wiki/%E5%8A%A0%E6%B1%80%E7%B4%8D%E7%AC%A6%E8%99%9F%E8%A1%A8" \o "加汀纳符号表)，涵盖1071个符号），[亚拉姆语](https://zh.wikipedia.org/wiki/%E4%BA%9A%E6%8B%89%E5%A7%86%E8%AF%AD" \o "亚拉姆语)，[巴拉维碑铭体](https://zh.wikipedia.org/w/index.php?title=%E5%B7%B4%E6%8B%89%E7%BB%B4%E7%A2%91%E9%93%AD%E4%BD%93&action=edit&redlink=1" \o "巴拉维碑铭体（页面不存在）)，[帕提亚碑铭体](https://zh.wikipedia.org/w/index.php?title=%E5%B8%95%E6%8F%90%E4%BA%9A%E7%A2%91%E9%93%AD%E4%BD%93&action=edit&redlink=1" \o "帕提亚碑铭体（页面不存在）)，[爪哇语](https://zh.wikipedia.org/wiki/%E7%88%AA%E5%93%87%E8%AF%AD" \o "爪哇语)，[凯提文](https://zh.wikipedia.org/wiki/%E5%87%B1%E6%8F%90%E6%96%87" \o "凯提文)，[傈僳族字母](https://zh.wikipedia.org/w/index.php?title=%E5%82%88%E5%83%B3%E6%97%8F%E5%AD%97%E6%AF%8D&action=edit&redlink=1" \o "傈僳族字母（页面不存在）)，[曼尼普尔文](https://zh.wikipedia.org/w/index.php?title=%E6%9B%BC%E5%B0%BC%E6%99%AE%E5%B0%94%E6%96%87&action=edit&redlink=1" \o "曼尼普尔文（页面不存在）)，[古南阿拉伯字母](https://zh.wikipedia.org/w/index.php?title=%E5%8F%A4%E5%8D%97%E9%98%BF%E6%8B%89%E4%BC%AF%E5%AD%97%E6%AF%8D&action=edit&redlink=1" \o "古南阿拉伯字母（页面不存在）)，[古突厥语](https://zh.wikipedia.org/wiki/%E5%8F%A4%E7%AA%81%E5%8E%A5%E8%AF%AD" \o "古突厥语)，[撒玛利亚语](https://zh.wikipedia.org/w/index.php?title=%E6%92%92%E7%8E%9B%E5%88%A9%E4%BA%9A%E8%AF%AD&action=edit&redlink=1" \o "撒玛利亚语（页面不存在）)，[大潭文](https://zh.wikipedia.org/w/index.php?title=%E5%A4%A7%E6%BD%AD%E6%96%87&action=edit&redlink=1" \o "大潭文（页面不存在）)和[大越文](https://zh.wikipedia.org/w/index.php?title=%E5%A4%A7%E8%B6%8A%E6%96%87&action=edit&redlink=1" \o "大越文（页面不存在）)。追加4,149个[中日韩统一表意文字](https://zh.wikipedia.org/wiki/%E4%B8%AD%E6%97%A5%E9%9F%A9%E7%BB%9F%E4%B8%80%E8%A1%A8%E6%84%8F%E6%96%87%E5%AD%97" \o "中日韩统一表意文字)（CJK-C），同时扩展了古[韩语](https://zh.wikipedia.org/wiki/%E9%9F%A9%E8%AF%AD" \o "韩语)和[吠陀梵语](https://zh.wikipedia.org/wiki/%E5%90%A0%E9%99%80%E6%A2%B5%E8%AA%9E" \o "吠陀梵语)的字符。[[16]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-17) |
| 6.0 | 2010年10月 |  | ISO/IEC 10646:2010与[印度卢比符号](https://zh.wikipedia.org/w/index.php?title=%E5%8D%B0%E5%BA%A6%E7%9B%A7%E6%AF%94%E7%AC%A6%E8%99%9F&action=edit&redlink=1)（英语：[Indian rupee sign](https://en.wikipedia.org/wiki/Indian_rupee_sign" \o "en:Indian rupee sign)） | 93 | 109,449 | 新增[巴塔克字母](https://zh.wikipedia.org/w/index.php?title=%E5%B7%B4%E5%A1%94%E5%85%8B%E5%AD%97%E6%AF%8D&action=edit&redlink=1" \o "巴塔克字母（页面不存在）)，[婆罗米文字](https://zh.wikipedia.org/wiki/%E5%A9%86%E7%BD%97%E7%B1%B3%E6%96%87%E5%AD%97" \o "婆罗米文字)，[曼达字母](https://zh.wikipedia.org/w/index.php?title=%E6%9B%BC%E8%BE%BE%E5%AD%97%E6%AF%8D&action=edit&redlink=1" \o "曼达字母（页面不存在）)，[纸牌](https://zh.wikipedia.org/wiki/%E7%BA%B8%E7%89%8C" \o "纸牌)符号，[交通标志](https://zh.wikipedia.org/wiki/%E4%BA%A4%E9%80%9A%E6%A0%87%E5%BF%97" \o "交通标志)，[地图](https://zh.wikipedia.org/wiki/%E5%9C%B0%E5%9B%BE" \o "地图)符号，[炼金符号](https://zh.wikipedia.org/w/index.php?title=%E7%82%BC%E9%87%91%E7%AC%A6%E5%8F%B7&action=edit&redlink=1" \o "炼金符号（页面不存在）)，[颜文字](https://zh.wikipedia.org/wiki/%E9%A2%9C%E6%96%87%E5%AD%97" \o "颜文字)和[绘文字](https://zh.wikipedia.org/wiki/%E7%B9%AA%E6%96%87%E5%AD%97" \o "绘文字)。追加222个额外的[中日韩统一表意文字](https://zh.wikipedia.org/wiki/%E4%B8%AD%E6%97%A5%E9%9F%A9%E7%BB%9F%E4%B8%80%E8%A1%A8%E6%84%8F%E6%96%87%E5%AD%97" \o "中日韩统一表意文字)（CJK-D）。[[17]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-18) |
| 6.1 | 2012年1月 |  | ISO/IEC 10646:2012 | 100 | 110,181 | [查克马字母](https://zh.wikipedia.org/w/index.php?title=%E6%9F%A5%E5%85%8B%E9%A9%AC%E5%AD%97%E6%AF%8D&action=edit&redlink=1" \o "查克马字母（页面不存在）), [麦罗埃文](https://zh.wikipedia.org/wiki/%E9%BA%A6%E7%BD%97%E5%9F%83%E6%96%87" \o "麦罗埃文), [麦罗埃象形文字](https://zh.wikipedia.org/w/index.php?title=%E9%BA%A6%E7%BD%97%E5%9F%83%E8%B1%A1%E5%BD%A2%E6%96%87%E5%AD%97&action=edit&redlink=1" \o "麦罗埃象形文字（页面不存在）), [波拉德文字](https://zh.wikipedia.org/w/index.php?title=%E6%B3%A2%E6%8B%89%E5%BE%B7%E6%96%87%E5%AD%97&action=edit&redlink=1" \o "波拉德文字（页面不存在）), [夏拉达文](https://zh.wikipedia.org/wiki/%E5%A4%8F%E6%8B%89%E9%81%94%E6%96%87" \o "夏拉达文), [索拉僧平文字](https://zh.wikipedia.org/wiki/%E7%B4%A2%E6%8B%89%E5%83%A7%E5%B9%B3%E6%96%87%E5%AD%97" \o "索拉僧平文字), and [泰克里文](https://zh.wikipedia.org/w/index.php?title=%E6%B3%B0%E5%85%8B%E9%87%8C%E6%96%87&action=edit&redlink=1" \o "泰克里文（页面不存在）).[[18]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-19) |
| 6.2 | 2012年9月 |  | ISO/IEC 10646:2012与[土耳其里拉符号](https://zh.wikipedia.org/wiki/%E5%9C%9F%E8%80%B3%E5%85%B6%E9%87%8C%E6%8B%89%E7%AC%A6%E5%8F%B7" \o "土耳其里拉符号) | 100 | 110,182 | [土耳其里拉符号](https://zh.wikipedia.org/wiki/%E5%9C%9F%E8%80%B3%E5%85%B6%E9%87%8C%E6%8B%89%E7%AC%A6%E5%8F%B7" \o "土耳其里拉符号)[[19]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-20)。 |
| 6.3 | 2013年9月 |  | ISO/IEC 10646:2012与6个字元 | 100 | 110,187 | 5个双向排版符号。[[20]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-21) |
| 7.0 | 2014年6月 |  | ISO/IEC 10646:2012与其第1、2修订版，以及[俄罗斯卢布](https://zh.wikipedia.org/wiki/%E4%BF%84%E7%BE%85%E6%96%AF%E7%9B%A7%E5%B8%83" \o "俄罗斯卢布)符号 | 123 | 113,021 | 新增[巴萨字母](https://zh.wikipedia.org/w/index.php?title=%E5%B7%B4%E8%90%A8%E5%AD%97%E6%AF%8D&action=edit&redlink=1" \o "巴萨字母（页面不存在）)，[高加索阿尔巴尼亚字母](https://zh.wikipedia.org/w/index.php?title=%E9%AB%98%E5%8A%A0%E7%B4%A2%E9%98%BF%E5%B0%94%E5%B7%B4%E5%B0%BC%E4%BA%9A%E5%AD%97%E6%AF%8D&action=edit&redlink=1" \o "高加索阿尔巴尼亚字母（页面不存在）)，[杜普雷严速记](https://zh.wikipedia.org/w/index.php?title=%E6%9D%9C%E6%99%AE%E9%9B%B7%E5%9A%B4%E9%80%9F%E8%A8%98&action=edit&redlink=1" \o "杜普雷严速记（页面不存在）)，[爱尔巴桑字母](https://zh.wikipedia.org/w/index.php?title=%E7%88%B1%E5%B0%94%E5%B7%B4%E6%A1%91%E5%AD%97%E6%AF%8D&action=edit&redlink=1" \o "爱尔巴桑字母（页面不存在）)，[帕拉瓦文](https://zh.wikipedia.org/wiki/%E5%B8%95%E6%8B%89%E7%93%A6%E6%96%87" \o "帕拉瓦文)，[可吉文](https://zh.wikipedia.org/w/index.php?title=%E5%8F%AF%E5%90%89%E6%96%87&action=edit&redlink=1" \o "可吉文（页面不存在）)，[库达瓦迪文](https://zh.wikipedia.org/w/index.php?title=%E5%BA%93%E8%BE%BE%E7%93%A6%E8%BF%AA%E6%96%87&action=edit&redlink=1" \o "库达瓦迪文（页面不存在）)，[线形文字A](https://zh.wikipedia.org/wiki/%E7%BA%BF%E5%BD%A2%E6%96%87%E5%AD%97A" \o "线形文字A)，[马哈佳尼文](https://zh.wikipedia.org/w/index.php?title=%E9%A9%AC%E5%93%88%E4%BD%B3%E5%B0%BC%E6%96%87&action=edit&redlink=1" \o "马哈佳尼文（页面不存在）)，[摩尼教字母](https://zh.wikipedia.org/wiki/%E6%91%A9%E5%B0%BC%E6%95%99%E5%AD%97%E6%AF%8D" \o "摩尼教字母)，[门得文字](https://zh.wikipedia.org/w/index.php?title=%E9%97%A8%E5%BE%97%E6%96%87%E5%AD%97&action=edit&redlink=1" \o "门得文字（页面不存在）)，[莫迪字母](https://zh.wikipedia.org/w/index.php?title=%E8%8E%AB%E8%BF%AA%E5%AD%97%E6%AF%8D&action=edit&redlink=1" \o "莫迪字母（页面不存在）)，[默文](https://zh.wikipedia.org/w/index.php?title=%E9%BB%98%E6%96%87&action=edit&redlink=1" \o "默文（页面不存在）)，[纳巴泰字母](https://zh.wikipedia.org/wiki/%E7%B4%8D%E5%B7%B4%E6%B3%B0%E5%AD%97%E6%AF%8D" \o "纳巴泰字母)，[古北阿拉伯文](https://zh.wikipedia.org/w/index.php?title=%E5%8F%A4%E5%8C%97%E9%98%BF%E6%8B%89%E4%BC%AF%E6%96%87&action=edit&redlink=1" \o "古北阿拉伯文（页面不存在）)，[古彼尔姆语字母](https://zh.wikipedia.org/w/index.php?title=%E5%8F%A4%E5%BD%BC%E5%B0%94%E5%A7%86%E8%AF%AD%E5%AD%97%E6%AF%8D&action=edit&redlink=1" \o "古彼尔姆语字母（页面不存在）)，[杨松录苗文](https://zh.wikipedia.org/w/index.php?title=%E6%9D%A8%E6%9D%BE%E5%BD%95%E8%8B%97%E6%96%87&action=edit&redlink=1" \o "杨松录苗文（页面不存在）)，[帕米拉文字](https://zh.wikipedia.org/w/index.php?title=%E5%B8%95%E7%B1%B3%E6%8B%89%E6%96%87%E5%AD%97&action=edit&redlink=1" \o "帕米拉文字（页面不存在）)，[袍清豪文](https://zh.wikipedia.org/w/index.php?title=%E8%A2%8D%E6%B8%85%E8%B1%AA%E6%96%87&action=edit&redlink=1" \o "袍清豪文（页面不存在）)，[诗篇巴列维文](https://zh.wikipedia.org/w/index.php?title=%E8%AF%97%E7%AF%87%E5%B7%B4%E5%88%97%E7%BB%B4%E6%96%87&action=edit&redlink=1" \o "诗篇巴列维文（页面不存在）)，[悉昙文字](https://zh.wikipedia.org/wiki/%E6%82%89%E6%9B%87%E6%96%87%E5%AD%97" \o "悉昙文字)，[提尔胡塔文](https://zh.wikipedia.org/w/index.php?title=%E6%8F%90%E5%B0%94%E8%83%A1%E5%A1%94%E6%96%87&action=edit&redlink=1" \o "提尔胡塔文（页面不存在）)，[瓦兰齐地文](https://zh.wikipedia.org/w/index.php?title=%E7%93%A6%E5%85%B0%E9%BD%90%E5%9C%B0%E6%96%87&action=edit&redlink=1" \o "瓦兰齐地文（页面不存在）)以及[装饰符号](https://zh.wikipedia.org/wiki/Dingbat" \o "Dingbat)。[[21]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-22) |
| 8.0 | 2015年6月 |  | ISO/IEC 10646:2014与其第1修订版，以及[乔治亚拉里](https://zh.wikipedia.org/wiki/%E5%96%AC%E6%B2%BB%E4%BA%9E%E6%8B%89%E9%87%8C" \o "乔治亚拉里)符号、9个中日韩统一表意文字与41个表情符号[[22]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-23) | 129 | 120,737 | 增加[阿洪姆文](https://zh.wikipedia.org/wiki/%E9%98%BF%E6%B4%AA%E5%A7%86%E6%96%87" \o "阿洪姆文)，[安纳托利亚象形文字](https://zh.wikipedia.org/w/index.php?title=%E5%AE%89%E7%BA%B3%E6%89%98%E5%88%A9%E4%BA%9A%E8%B1%A1%E5%BD%A2%E6%96%87%E5%AD%97&action=edit&redlink=1" \o "安纳托利亚象形文字（页面不存在）)，[哈坦文](https://zh.wikipedia.org/w/index.php?title=%E5%93%88%E5%9D%A6%E6%96%87&action=edit&redlink=1" \o "哈坦文（页面不存在）)，[穆尔塔尼文](https://zh.wikipedia.org/w/index.php?title=%E7%A9%86%E5%B0%94%E5%A1%94%E5%B0%BC%E6%96%87&action=edit&redlink=1" \o "穆尔塔尼文（页面不存在）)，[古匈牙利字母](https://zh.wikipedia.org/wiki/%E5%8F%A4%E5%8C%88%E7%89%99%E5%88%A9%E5%AD%97%E6%AF%8D" \o "古匈牙利字母)，[书写符号](https://zh.wikipedia.org/w/index.php?title=%E4%B9%A6%E5%86%99%E7%AC%A6%E5%8F%B7&action=edit&redlink=1" \o "书写符号（页面不存在）)，5,771个[中日韩统一表意文字](https://zh.wikipedia.org/wiki/%E4%B8%AD%E6%97%A5%E9%9F%A9%E7%BB%9F%E4%B8%80%E8%A1%A8%E6%84%8F%E6%96%87%E5%AD%97" \o "中日韩统一表意文字)字符，[切罗基语](https://zh.wikipedia.org/wiki/%E5%88%87%E7%BD%97%E5%9F%BA%E8%AF%AD" \o "切罗基语)小写字母，以及五种[绘文字](https://zh.wikipedia.org/wiki/%E7%B9%AA%E6%96%87%E5%AD%97" \o "绘文字)[肤色](https://zh.wikipedia.org/wiki/%E8%86%9A%E8%89%B2" \o "肤色)修改字符。[[23]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-24) |
| 9.0 | 2016年6月 |  | ISO/IEC 10646:2014与其第1、2修订版，阿德拉姆字母、尼泊尔纽瓦字母、日本电视符号和74个绘文字表情与符号。[[24]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-25) | 135 | 128,237 | 新增[阿德拉姆字母](https://zh.wikipedia.org/w/index.php?title=%E9%98%BF%E5%BE%B7%E6%8B%89%E5%A7%86%E5%AD%97%E6%AF%8D&action=edit&redlink=1" \o "阿德拉姆字母（页面不存在）)，[比奇舒奇文](https://zh.wikipedia.org/w/index.php?title=%E6%AF%94%E5%A5%87%E8%88%92%E5%A5%87%E6%96%87&action=edit&redlink=1" \o "比奇舒奇文（页面不存在）)，[象雄文](https://zh.wikipedia.org/wiki/%E8%B1%A1%E9%9B%84%E6%96%87" \o "象雄文)，[尼泊尔纽瓦字母](https://zh.wikipedia.org/w/index.php?title=%E5%B0%BC%E6%B3%8A%E5%B0%94%E7%BA%BD%E7%93%A6%E5%AD%97%E6%AF%8D&action=edit&redlink=1" \o "尼泊尔纽瓦字母（页面不存在）)，[欧塞奇字母](https://zh.wikipedia.org/w/index.php?title=%E6%AC%A7%E5%A1%9E%E5%A5%87%E5%AD%97%E6%AF%8D&action=edit&redlink=1" \o "欧塞奇字母（页面不存在）)，[西夏文](https://zh.wikipedia.org/wiki/%E8%A5%BF%E5%A4%8F%E6%96%87" \o "西夏文)以及74个绘文字[[25]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-26) |
| 10.0 | 2017年6月 |  | ISO/IEC 10646:2017，新增56个[绘文字](https://zh.wikipedia.org/wiki/%E7%B9%AA%E6%96%87%E5%AD%97" \o "绘文字)符号，385个[变体假名](https://zh.wikipedia.org/wiki/%E8%AE%8A%E9%AB%94%E5%81%87%E5%90%8D" \o "变体假名)字符，和3个札那巴札尔字符[[2]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-Unicode10.0-2) | 139 | 136,755 | [札那巴札尔](https://zh.wikipedia.org/wiki/%E8%92%99%E5%8F%A4%E6%96%87%E5%AD%97" \o "蒙古文字)、[索永布文字](https://zh.wikipedia.org/wiki/%E7%B4%A2%E6%B0%B8%E5%B8%83%E6%96%87%E5%AD%97" \o "索永布文字)、[马萨拉姆共地文字](https://zh.wikipedia.org/w/index.php?title=%E9%A9%AC%E8%90%A8%E6%8B%89%E5%A7%86%E5%85%B1%E5%9C%B0%E6%96%87%E5%AD%97&action=edit&redlink=1" \o "马萨拉姆共地文字（页面不存在）)、[女书](https://zh.wikipedia.org/wiki/%E5%A5%B3%E4%B9%A6" \o "女书)、[变体假名](https://zh.wikipedia.org/wiki/%E8%AE%8A%E9%AB%94%E5%81%87%E5%90%8D" \o "变体假名)（非标准[平假名](https://zh.wikipedia.org/wiki/%E5%B9%B3%E5%81%87%E5%90%8D" \o "平假名)），7,494个[中日韩统一表意文字](https://zh.wikipedia.org/wiki/%E4%B8%AD%E6%97%A5%E9%9F%93%E7%B5%B1%E4%B8%80%E8%A1%A8%E6%84%8F%E6%96%87%E5%AD%97" \o "中日韩统一表意文字)与56个[绘文字](https://zh.wikipedia.org/wiki/%E7%B9%AA%E6%96%87%E5%AD%97" \o "绘文字)[[26]](https://zh.wikipedia.org/wiki/Unicode" \l "cite_note-27) |

### 编码方式

统一码的编码方式与[ISO 10646](https://zh.wikipedia.org/wiki/ISO_10646" \o "ISO 10646)的[通用字符集](https://zh.wikipedia.org/wiki/%E9%80%9A%E7%94%A8%E5%AD%97%E7%AC%A6%E9%9B%86" \o "通用字符集)概念相对应。目前实际应用的统一码版本对应于[UCS-2](https://zh.wikipedia.org/wiki/UCS-2" \o "UCS-2)，使用16[位](https://zh.wikipedia.org/wiki/%E4%BD%8D%E5%85%83" \o "位元)的编码空间。也就是每个字符占用2个[字节](https://zh.wikipedia.org/wiki/%E5%AD%97%E8%8A%82" \o "字节)。这样理论上一共最多可以表示216（即65536）个字符。基本满足各种语言的使用。实际上目前版本的统一码并未完全使用这16位编码，而是保留了大量空间以作为特殊使用或将来扩展。

上述16位统一码字符构成[基本多文种平面](https://zh.wikipedia.org/wiki/%E5%9F%BA%E6%9C%AC%E5%A4%9A%E6%96%87%E7%A8%AE%E5%B9%B3%E9%9D%A2" \o "基本多文种平面)。最新（但未实际广泛使用）的统一码版本定义了16个[辅助平面](https://zh.wikipedia.org/wiki/%E8%BE%85%E5%8A%A9%E5%B9%B3%E9%9D%A2" \o "辅助平面)，两者合起来至少需要占据21位的编码空间，比3字节略少。但事实上辅助平面字符仍然占用4字节编码空间，与[UCS-4](https://zh.wikipedia.org/wiki/UCS-4" \o "UCS-4)保持一致。未来版本会扩充到ISO 10646-1实现级别3，即涵盖UCS-4的所有字符。UCS-4是一个更大的尚未填充完全的31位字符集，加上恒为0的首位，共需占据32位，即4字节。理论上最多能表示231个字符，完全可以涵盖一切语言所用的符号。

基本多文种平面的字符的编码为*U+hhhh*，其中每个*h*代表一个[十六进制](https://zh.wikipedia.org/wiki/%E5%8D%81%E5%85%AD%E8%BF%9B%E5%88%B6" \o "十六进制)数字，与UCS-2编码完全相同。而其对应的4字节UCS-4编码后两个字节一致，前两个字节则所有位均为0。

关于统一码和ISO 10646及UCS的详细关系，见[通用字符集](https://zh.wikipedia.org/wiki/%E9%80%9A%E7%94%A8%E5%AD%97%E7%AC%A6%E9%9B%86" \o "通用字符集)。

### 实现方式

Unicode的实现方式不同于编码方式。一个字符的Unicode编码是确定的。但是在实际传输过程中，由于不同[系统平台](https://zh.wikipedia.org/wiki/%E7%B3%BB%E7%BB%9F%E5%B9%B3%E5%8F%B0" \o "系统平台)的设计不一定一致，以及出于节省空间的目的，对Unicode编码的实现方式有所不同。Unicode的实现方式称为**Unicode转换格式**（Unicode Transformation Format，简称为UTF）

例如，如果一个仅包含基本7位[ASCII](https://zh.wikipedia.org/wiki/ASCII" \o "ASCII)字符的Unicode文件，如果每个字符都使用2字节的原Unicode编码传输，其第一字节的8位始终为0。这就造成了比较大的浪费。对于这种情况，可以使用UTF-8编码，这是一种变长编码，它将基本7位ASCII字符仍用7位编码表示，占用一个字节（首位补0）。而遇到与其他Unicode字符混合的情况，将按一定算法转换，每个字符使用1-3个字节编码，并利用首位为0或1进行识别。这样对以7位ASCII字符为主的西文文档就大幅节省了编码长度（具体方案参见*[UTF-8](https://zh.wikipedia.org/wiki/UTF-8" \o "UTF-8)*）。类似的，对未来会出现的需要4个字节的辅助平面字符和其他UCS-4扩充字符，2字节编码的[UTF-16](https://zh.wikipedia.org/wiki/UTF-16" \o "UTF-16)也需要通过一定的算法进行转换。

再如，如果直接使用与Unicode编码一致（仅限于BMP字符）的UTF-16编码，由于每个字符占用了两个字节，在[麦金塔电脑](https://zh.wikipedia.org/wiki/%E9%BA%A5%E9%87%91%E5%A1%94%E9%9B%BB%E8%85%A6" \o "麦金塔电脑)（[Mac](https://zh.wikipedia.org/wiki/Mac" \o "Mac)）机和[个人电脑](https://zh.wikipedia.org/wiki/%E5%80%8B%E4%BA%BA%E9%9B%BB%E8%85%A6" \o "个人电脑)上，对字节顺序的理解是不一致的。这时同一字节流可能会被解释为不同内容，如某字符为十六进制编码4E59，按两个字节拆分为4E和59，在Mac上读取时是从低字节开始，那么在Mac OS会认为此4E59编码为594E，找到的字符为“奎”，而在Windows上从高字节开始读取，则编码为U+4E59的字符为“乙”。就是说在Windows下以UTF-16编码保存一个字符“乙”，在Mac OS环境下开启会显示成“奎”。此类情况说明UTF-16的编码顺序若不加以人为定义就可能发生混淆，于是在UTF-16编码实现方式中使用了[大端序](https://zh.wikipedia.org/wiki/%E5%AD%97%E8%8A%82%E5%BA%8F" \o "字节序)（Big-Endian，简写为UTF-16 BE）、[小端序](https://zh.wikipedia.org/wiki/%E5%AD%97%E8%8A%82%E5%BA%8F" \o "字节序)（Little-Endian，简写为UTF-16 LE）的概念，以及可附加的[位元组顺序记号](https://zh.wikipedia.org/wiki/%E4%BD%8D%E5%85%83%E7%B5%84%E9%A0%86%E5%BA%8F%E8%A8%98%E8%99%9F" \o "位元组顺序记号)解决方案，目前在PC机上的Windows系统和Linux系统对于UTF-16编码默认使用UTF-16 LE。（具体方案参见*[UTF-16](https://zh.wikipedia.org/wiki/UTF-16" \o "UTF-16)*）

此外Unicode的实现方式还包括[UTF-7](https://zh.wikipedia.org/wiki/UTF-7" \o "UTF-7)、[Punycode](https://zh.wikipedia.org/wiki/Punycode" \o "Punycode)、[CESU-8](https://zh.wikipedia.org/w/index.php?title=CESU-8&action=edit&redlink=1" \o "CESU-8（页面不存在）)、[SCSU](https://zh.wikipedia.org/w/index.php?title=SCSU&action=edit&redlink=1" \o "SCSU（页面不存在）)、[UTF-32](https://zh.wikipedia.org/wiki/UTF-32" \o "UTF-32)、[GB18030](https://zh.wikipedia.org/wiki/GB18030" \o "GB18030)等，这些实现方式有些仅在一定的国家和地区使用，有些则属于未来的规划方式。目前通用的实现方式是UTF-16小端序（LE）、UTF-16大端序（BE）和UTF-8。在微软公司[Windows XP](https://zh.wikipedia.org/wiki/Windows_XP" \o "Windows XP)附带的[记事本](https://zh.wikipedia.org/wiki/%E8%AE%B0%E4%BA%8B%E6%9C%AC" \o "记事本)（Notepad）中，“另存为”对话框可以选择的四种编码方式除去非Unicode编码的[ANSI](https://zh.wikipedia.org/wiki/ANSI" \o "ANSI)（对于英文系统即[ASCII](https://zh.wikipedia.org/wiki/ASCII" \o "ASCII)编码，中文系统则为[GB2312](https://zh.wikipedia.org/wiki/GB2312" \o "GB2312)或[Big5](https://zh.wikipedia.org/wiki/Big5" \o "Big5)编码）外，其余三种为“Unicode”（对应UTF-16 LE）、“Unicode big endian”（对应UTF-16 BE）和“UTF-8”。

目前辅助平面的工作主要集中在第二和第三平面的[中日韩统一表意文字](https://zh.wikipedia.org/wiki/%E4%B8%AD%E6%97%A5%E9%9F%A9%E7%BB%9F%E4%B8%80%E8%A1%A8%E6%84%8F%E6%96%87%E5%AD%97" \o "中日韩统一表意文字)中，因此包括[GBK](https://zh.wikipedia.org/wiki/GBK" \o "GBK)、[GB18030](https://zh.wikipedia.org/wiki/GB18030" \o "GB18030)、[Big5](https://zh.wikipedia.org/wiki/Big5" \o "Big5)等[简体中文](https://zh.wikipedia.org/wiki/%E7%AE%80%E4%BD%93%E4%B8%AD%E6%96%87" \o "简体中文)、[繁体中文](https://zh.wikipedia.org/wiki/%E7%B9%81%E4%BD%93%E4%B8%AD%E6%96%87" \o "繁体中文)、[日文](https://zh.wikipedia.org/wiki/%E6%97%A5%E6%96%87" \o "日文)、[韩文](https://zh.wikipedia.org/wiki/%E9%9F%A9%E6%96%87" \o "韩文)以及越南[喃字](https://zh.wikipedia.org/wiki/%E5%96%83%E5%AD%97" \o "喃字)的各种编码与Unicode的协调性被重点关注。考虑到Unicode最终要涵盖所有的字符。从某种意义而言，这些编码方式也可视作Unicode的出现于其之前的*既成事实*的实现方式，如同[ASCII](https://zh.wikipedia.org/wiki/ASCII" \o "ASCII)及其扩展[Latin-1](https://zh.wikipedia.org/wiki/Latin-1" \o "Latin-1)一样，后两者的字符在16位Unicode编码空间中的编码第一字节各位全为0，第二字节编码与原编码完全一致。但上述东亚语言编码与Unicode编码的对应关系要复杂得多。

### Unicode字符平面映射

## 非Unicode环境

在非Unicode环境下，由于不同国家和地区采用的字符集不一致，很可能出现无法正常显示所有字符的情况。微软公司使用了[代码页](https://zh.wikipedia.org/wiki/%E4%BB%A3%E7%A0%81%E9%A1%B5" \o "代码页)（Codepage）转换表的技术来过渡性的部分解决这一问题，即通过指定的转换表将非Unicode的字符编码转换为同一字符对应的系统内部使用的Unicode编码。可以在“语言与区域设置”中选择一个代码页作为非Unicode编码所采用的默认编码方式，如936为简体中文[GB码](https://zh.wikipedia.org/wiki/%E5%9B%BD%E5%AE%B6%E6%A0%87%E5%87%86%E4%BB%A3%E7%A0%81" \o "国家标准代码)，950为繁体中文[Big5](https://zh.wikipedia.org/wiki/Big5" \o "Big5)（皆指PC上使用的）。在这种情况下，一些非英语的欧洲语言编写的软件和文档很可能出现乱码。而将代码页设置为相应语言中文处理又会出现问题，这一情况无法避免。只有完全采用统一编码才能彻底解决这些问题，但目前尚无法做到这一点。

代码页技术现在广泛为各种平台所采用。UTF-7的代码页是65000，UTF-8的代码页是65001。

引自维基百科unicode字条，网址：https://zh.wikipedia.org/wiki/Unicode

以上内容引自维基百科GB2312,BIG5,UTF-16,Unicode字条，网址：https://zh.wikipedia.org/wiki/Wikipedia:%E9%A6%96%E9%A1%B5