

表16-1 MIME charset 编码标记

MIME charset 值	描 述
us-ascii	这是个著名的字符编码，在 1968 年就已标准化，称为 ANSI_X3.4-1968。它也称为 ASCII，但最好还是加上“US”前缀，因为 ISO 646 中有某些国际化的变体，它们修改了一些字符。US-ASCII 把 7 位数值映射到 128 个字符上。最高位未使用
iso-8859-1	iso-8859-1 是对 ASCII 的 8 位扩展，以支持西欧的多种语言。它使用了最高位以包含更多西欧字符，而保持 ASCII 的编码部分（0 ~ 127）没有变。它也称为 iso-latin-1，或简称为 Latin1
iso-8859-2	对 ASCII 扩展以包括中欧和东欧语言中的字符，包括捷克、波兰、罗马尼亚。它也称为 iso-latin-2
iso-8859-5	对 ASCII 扩展以包括斯拉夫语字符，使用这些字符的语言包括俄语、塞尔维亚语和保加利亚语
iso-8859-6	对 ASCII 扩展以包括阿拉伯语字符。因为阿拉伯语字符的显示形状会随它在单词中的位置而变化，阿拉伯语的显示引擎需要分析上下文来为每个字符生成正确的形状
iso-8859-7	对 ASCII 扩展以包括现代希腊语字符。以前称为 ELOT-928 或 ECMA-118:1986
iso-8859-8	对 ASCII 扩展以包括希伯来语和意第绪语（这两种语言都是犹太人所用的）的字符
iso-8859-15	更新了 iso-8859-1，用遗漏的法语和芬兰语字母替换了一些不太常用的标点符号和分数符号，并用新的欧元符号替换国际货币符号。这种字符集简称为 Latin0，可能将来会替代 iso-8859-1，作为欧洲的首选默认字符集
iso-2022-jp	iso-2022-jp 是在日语的电子邮件和网页内容中广泛使用的编码。它是一种变长编码方案，支持用单字节表示 ASCII 字符，但使用 3 字符的模式转义序列在 3 种日语字符集中切换
euc-jp	euc-jp 是遵循 ISO 2022 的变长编码，它用显式的二进制码模式来标识每个字符，不需要模式及转义序列。它使用单字节、2 字节以及 3 字节的序列来标识多个日语字符集中的字符
Shift_JIS	该编码起初是由微软公司开发的，有时称为 SJIS 或 MS Kanji。出于保持历史兼容性方面的原因，它有点儿复杂，并且不能映射所有的字符，不过它还是用的很普遍
koi8-r	KOI8-R 是为俄语设计的流行的 8 位因特网字符集编码，在 IETF RFC 1489 中定义。这些大写字母是 Code for Information Exchange, 8 bit, Russian（俄语 8 位信息交换代码）的首字母缩略形式
utf-8	UTF-8 是一种用来表示 UCS（Unicode）的常用变长字符编码方案，UCS 的意思是 Universal Character Set of the world's characters（世界字符统一字符集）。UTF-8 使用变长的编码来表示字符代码值，每个字符使用 1 ~ 6 个字节。UTF-8 的主要特点之一就是保持对普通的 7 位 ASCII 文本的后向兼容性
windows-1252	微软公司把它编码后的字符集称为 code page（代码页）。Windows 的代码页 1252（也称为 CP1252 或 WinLatin1）是对 iso-8859-1 的扩展