

1. US-ASCII:所有字符集的始祖

379

ASCII 是最著名的编码后字符集，早在 1968 年就被 ANSI 在标准 X3.4，“美国标准信息交换代码”（American Standard Code for Information Interchange）中进行了标准化。ASCII 的代码值只是从 0 到 127，因此只需要 7 个二进制码就可以覆盖代码空间。ASCII 的推荐名称是 US-ASCII，这样可以和那些 7 位字符集的一些国际化变体区分开来。

HTTP 报文（首部、URI 等）使用的字符集是 US-ASCII。

2. iso-8859

iso-8859 字符集标准是 US-ASCII 的 8 位超集，使用二进制码的高位增加了一些国际化书面字符。由额外的二进制码提供的附加空间（多了 128 个代码）还不够大，甚至都不够所有的欧洲字符使用，更不用说亚洲字符了。因此 iso-8859 为不同地区定制了不同的字符集，如下所示。

iso-8859-1	西欧语言（例如，英语、法语）
iso-8859-2	中欧和东欧语言（例如，捷克、波兰）
iso-8859-3	南欧语言
iso-8859-4	北欧语言（例如，拉托维亚，立陶宛，格陵兰）
iso-8859-5	斯拉夫语（例如，保加利亚、俄罗斯、塞尔维亚）
iso-8859-6	阿拉伯语
iso-8859-7	希腊语
iso-8859-8	希伯来语
iso-8859-9	土耳其语
iso-8859-10	日耳曼和斯堪的纳维亚语言（例如，冰岛、因纽特）
iso-8859-15	对 iso-8859-1 的修改，包括了新的欧元字符

iso-8859-1 也称为 Latin1，是 HTML 的默认字符集。可以用它来表示大多数西欧语言的文本。因为 iso-8859-15 中包含了新的欧元符号，有过一些用它来代替 iso-8859-1 并作为 HTTP 默认编码后字符集的讨论。然而，由于 iso-8859-1 已经被广泛采用，要大范围地变更到 iso-8859-15 恐怕不是短时间内可以完成的。

3. JIS X 0201

JIS X 0201 是把 ASCII 扩展到日文半宽片假名字符的一个极小化的字符集。半宽片假名字符最早用在日文电报系统中。JIS X 0201 常常被称作 JIS Roman，JIS 是“Japanese Industrial Standard”（日文工业化标准）的缩写。

4. JIS X 0208与JIS X 0212

日文中包括数千个来自几个书面语系统中的字符。尽管可以（很痛苦地）勉强只使用 JIS X 0201 中的那 63 个基本的片假名字符，但实际使用中需要远比这个更完整的字符集。

380