381 下面我们来看一些常见的编码方案。

1. 8位

8 位固定宽度恒等编码把每个字符代码编码为相应的 8 位二进制值。它只能支持有 256 个字符代码范围的字符集。iso-8859 字符集家族系列使用的就是 8 位恒等编码。

2. UTF-8

UTF-8 是一种流行的为 UCS 设计的字符编码方案，UTF 表示 UCS 变换格式（UCS Transformation Format）。UTF-8 为字符代码值使用的是无模态的变宽编码，第一字节的高位表示编码后的字符所用的字节数，所需的每个后续字节都含有 6 位的代码值（参见表 16-2）。

如果编码后的第 1 字节的最高位是 0，长度就是 1 字节，剩余的 7 位就包含字符的代码。这样带来的美妙结果就是和 ASCII 兼容（但和 iso-8859 系列不兼容，因为 iso-8859 系列使用了最高位）。

表16-2 UTF-8 变宽无模态编码

字符代码的二进制位	字节1	字节2	字节3	字节4	字节5	字节 6
0-7	0ccccccc	-	-	-	-	-
8-11	110ccccc	10cccccc	-	-	-	-
12-16	1110cccc	10cccccc	10cccccc	-	-	-
17-21	11110ccc	10cccccc	10cccccc	10cccccc	-	-
22-26	111110cc	10cccccc	10cccccc	10cccccc	10cccccc	-
27-31	1111110c	10cccccc	10cccccc	10cccccc	10cccccc	10cccccc

例如，字符代码 90（ASCII 的“Z”）会被编码为 1 个字节（01011010），而代码 5073（13 位二进制值为 1001111010001）会被编码为 3 个字节：

11100001 10001111 10010001

3. iso-2022-jp

iso-2022-jp 是互联网上的日文文档中广泛使用的编码。它是变宽、有模态的，所有值都不超过 128，以避免和不支持 8 位字符的软件出现兼容性问题。

编码上下文始终被设置为 4 种预设的字符集之一[12]，使用特殊的“转义序列”（escape sequence）在字符集之间切换。iso-2022-jp 的初始状态使用 US-ASCII 字符

注 12：iso-2022-jp 编码和这 4 种字符集是紧密绑定的，而其他一些编码是和特定的字符集无关的。