

- 编码后的字符集

组成字符库（从全球的字符中选出若干字符）的已编码字符集，并为每个字符分配代码空间中的一个代码。换句话说，它把数字化的字符代码映射为实际的字符。

- 字符编码方案

把数字化的字符代码编码成一系列二进制码（并能相应地反向解码）的算法。字符编码方案可用来减少识别字符所需要的数据总量（压缩）、解决传输限制、统一重叠编码字符集。

16.3.2 字符集的命名很糟糕

从技术上说，MIME 中的 charset 标记（用在 Content-Type 首部的 charset 参数中和 Accept-Charset 首部中）描述的压根就不是字符集。MIME 中的 charset 值所命名的是把数据位映射为唯一的字符的一整套算法。它是字符编码方案（character encoding scheme）和编码后的字符集（coded character set）这两种概念的组合（参见图 16-2）。

因为关于字符编码方案和编码后的字符集方面的标准都已经发布过了，所以，这个术语的使用是很草率的，很容易引起混淆⁵。下面是 HTTP/1.1 的作者们对于他们如何使用这些术语的介绍（在 RFC2616 中）。

术语“字符集”在本文档中是指一种方法，它可以把一系列 8 位字节转换为一系列字符。注意：术语“字符集”经常被称为“字符编码”。但由于 HTTP 和 MIME 共享同样的注册信息，术语也能共享是很重要的。

IETF 在 RFC 2277 中也采用了非标准的术语。

本文档中使用术语“字符集”来表示一组把一系列 8 位字节转换为一系列字符的规则的组合，比如编码后的字符集与字符编码方案的组合。这与 MIME 的“charset=” 参数中标识符的用法相同，并且已在 IANA 的字符集注册表中注册。（注意这不是在其他标准主体，比如在国际标准化组织 ISO 中使用的术语）。

注 5：更糟糕的是，MIME 中的 charset 标记经常会从特定的编码后字符集的名称或编码方案的名称里面选取。例如，iso-8859-1 是一个编码后字符集（它为一个包含 256 个欧洲字符的集合分配了数字化的代码），但 MIME 用 charset 值 iso-8859-1 来表示一种 8 位的、对编码后的字符集恒等的编码。这种不精确的术语并不是致命的问题，但在阅读标准文档的时候，需要对其假设用法保持清醒的头脑。