

16.5.3 转义和反转义

URI 转义提供了一种安全的方式，可以在 URI 内部插入保留字符以及原本不支持的字符（比如各种空白）。每个转义是一组 3 字符序列，由百分号（%）后面跟上两个十六进制数字的字符。这两个十六进制数字就表示一个 US-ASCII 字符的代码。

例如，要在 URL 中插入一个空白（ASCII 32），可以用转义 %20，因为 20 是 32 的十六进制表示。类似地，如果想插入一个百分号并且不想让它被当作转义，就可以输入 %25，25 是百分号的 ASCII 代码的十六进制值。

图 16-10 展示了 URI 中的概念性字符是如何转换为当前字符集中字符的代码字节的。需要处理 URI 时，转义会被反转义回来，产生它们代表的 ASCII 代码的字节。

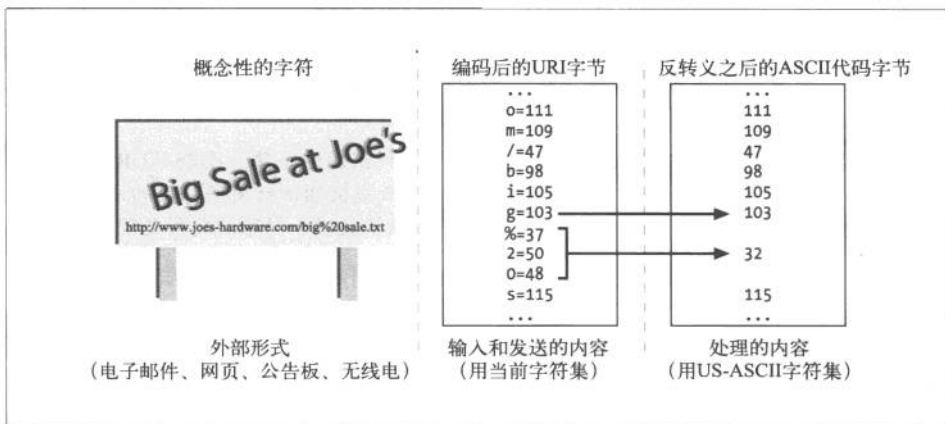


图 16-10 URI 中的字符在传输的时候要转义，但处理的时候要反转义

在内部处理时，HTTP 应用程序应当在传输和转发 URI 的时候保持转义不变。HTTP 应用程序应该仅在需要数据的时候才对 URI 进行转义。更重要的是，应用程序应该确保任何 URI 都不会被反转义 2 次，因为在转义的时候可能会把百分号编码进去，反转义出来之后，再转一次就会导致数据丢失。

16.5.4 转义国际化字符

需要注意的是，要转义的值本身应该在 US-ASCII 代码值的范围内（0 ~ 127）。某些应用程序试图用转义值来表示 iso-8859-1 中扩展的字符（代码范围在 128 ~ 255）。例如，网站服务器可能会错误地用转义来对包含了国际字符的文件名进行编码。这样做是不对的，可能会使别的应用出问题。

390