

JIS X 0208 字符集是首个多字节日文字符集，它定义了 6879 个编码的字符，其中大多数是来源于中文的日本汉字。JIS X 0212 字符集又扩充了 6067 个字符。

5. UCS

UCS (Universal Character Set, 统一字符集) 是把全世界的所有字符整合到单一的编码后字符集的环球标准化成果。UCS 由 ISO 10646 定义。Unicode 是遵循 UCS 标准的商业化联合组织。UCS 具有能容纳百万以上字符的代码空间，不过基本集合只有大约 5 万个字符。

16.3.6 字符编码方案

字符编码方案规定如何把字符的代码数字打包装入内容比特，以及在另一端如何将其解包回字符代码（参见图 16-7）。字符编码方案有以下 3 种主要类型。

- 固定宽度

固定宽度方式的编码用固定数量的比特表示每个编码后的字符。它们能被快速处理，但可能会浪费空间。

- 可变宽度（无模态）

可变宽度方式的编码对不同的字符代码数字采用不同数量的比特。对于常用字符，这样可以减少需要的位数，而且还能在允许使用多字节来表示国际性字符的同时，保持对传统 8 位字符集的兼容性。

- 可变宽度（有模态）

有模态的编码使用特殊的“转义”模式在不同的模态之间切换。例如，可以用有模态的编码在文本中使用多个互相有重叠的字符集。有模态的编码处理起来比较复杂，但它们可以有效地支持复杂的书写系统。

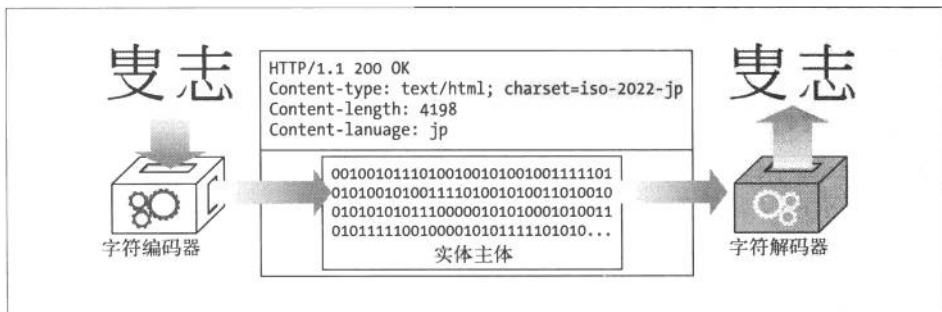


图 16-7 字符编码方案把字符代码编码为比特，并负责反向变换