

参数“q=0.8”是质量因子（quality factor），说明英语的优先级（0.8）比法语低（默认值是 1.0）。

## 16.2 字符集与HTTP

现在我们进入主题，开始研究网站国际化中最重要（且令人困惑）的方面——各国的字母表和它们的字符集编码。

Web 字符集标准很有些令人迷惑。由于必须阅读很多标准文档，其中术语复杂且不一致，再加上对外语不太熟悉，很多人首次尝试编写国际化的网站软件时，都被搞糊涂了。本节和下一节应该能让读者更容易地学会在 HTTP 中使用字符集。

### 16.2.1 字符集是把字符转换为二进制码的编码

HTTP 字符集的值说明如何把实体内容的二进制码转换为特定字母表中的字符。每个字符集标记都命名了一种把二进制码转换为字符的算法（反之亦然）。字符集标记在由 IANA 维护（参见 <http://www.iana.org/assignments/character-sets>）的 MIME 字符集注册机构进行了标准化。附录 H 中概述了其中的很多字符集。

下面的 Content-Type 首部告知接收者，传输的内容是一份 HTML 文件，用 charset 参数告知接收者使用 iso-8859-6 阿拉伯字符集的解码算法把内容中的二进制码转换为字符：

```
Content-Type: text/html; charset=iso-8859-6
```

iso-8859-6 的编码算法把 8 位值域映射为拉丁字母和阿拉伯字母，以及数字，标点和其他符号<sup>1</sup>。例如，在图 16-1 中，突出显示的二进制码的值是 225，它在 iso-8859-6 中被映射到阿拉伯字母“FEH”（读音类似英语字母 F）。

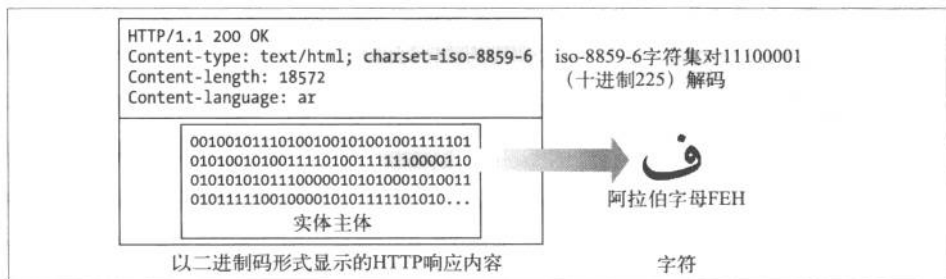


图 16-1 charset 参数告知客户端如何把内容中的二进制码转换为字符

注 1：与汉语、日语不同的是，阿拉伯语中只有 28 个字符。8 位空间有 256 个不同的值，足以容纳拉丁字符、阿拉伯字符以及其他符号。