

每天有上亿的人用数百种语言写着各种文档。为了真正实现万维网的目标，HTTP 要能够传输和处理用多种语言和字母表编写的国际性文档。

本章涵盖网站国际化方面的两个主要问题：字符集编码（character set encoding）和语言标记（language tag）。HTTP 应用程序使用字符集编码请求和显示不同字母表中的文本，它们使用语言标记根据用户所理解的语言来说明并限制内容。而在本章的最后将讨论多语言 URI 和日期格式。

本章主要内容：

- 讲解 HTTP 如何与多语言字母表的方案和相关标准进行交互；
- 快速概览术语、技术和标准，以帮助 HTTP 编程人员正确理解（熟悉字符编码的读者可以跳过本节）；
- 解释对各种语言的标准命名系统，以及标准化的语言标记如何描述和选择内容；
- 概述国际性的 URI 要遵循的规则和注意事项；
- 简要讨论日期格式和其他国际化方面的问题。

16.1 HTTP对国际性内容的支持

HTTP 报文中可以承载以任何语言表示的内容，就像它能承载图像、影片，或任何类型的媒体那样。对 HTTP 来说，实体主体只是二进制信息的容器而已。

370

为了支持国际性的内容，服务器需要告知客户端每个文档的字母表和语言，这样客户端才能正确地把文档中的信息解包为字符并把内容呈现给用户。

服务器通过 HTTP 协议的 Content-Type 首部中的 charset 参数和 Content-Language 首部告知客户端文档的字母表和语言。这些首部描述了实体主体的“信息盒子”里面装的是什麼，如何把内容转换成合适的字符以便显示在屏幕上以及里面的词语表示的是哪种语言。

同时，客户端需要告知服务器用户理解何种语言，浏览器上安装了何种字母表编码算法。客户端发送 Accept-Charset 首部和 Accept-Language 首部，告知服务器它理解哪些字符集编码算法和语言以及其中的优先顺序。

下面的 HTTP 报文中的这些 Accept 首部可能是母语为法语的人发出的。他喜欢使用母语，但也会说一点儿英语，他的浏览器支持 iso-8859-1 西欧字符集编码和 UTF-8 Unicode 字符集编码：

```
Accept-Language: fr, en;q=0.8
Accept-Charset: iso-8859-1, utf-8
```