

旁注 在连接中 EOF 意味着什么?

EOF 的概念常常使人们感到迷惑,尤其是在因特网连接的上下文中。首先,我们需要理解其实并没有像 EOF 字符这样的一个东西。进一步来说,EOF 是由内核检测到的一种条件。应用程序在它接收到一个由 read 函数返回的零返回码时,它就会发现出 EOF 条件。对于磁盘文件,当前文件位置超出文件长度时,会发生 EOF。对于因特网连接,当一个进程关闭连接它的那一端时,会发生 EOF。连接另一端的进程在试图读取流中最后一个字节之后的字节时,会检测到 EOF。

11.5 Web 服务器

迄今为止,我们已经在简单的 echo 服务器的上下文中讨论了网络编程。在这一节里,我们将向你展示如何利用网络编程的基本概念,来创建你自己的虽小但功能齐全的 Web 服务器。

11.5.1 Web 基础

Web 客户端和服务端之间的交互用的是一个基于文本的应用级协议,叫做 HTTP (Hypertext Transfer Protocol, 超文本传输协议)。HTTP 是一个简单的协议。一个 Web 客户端(即浏览器)打开一个到服务器的因特网连接,并且请求某些内容。服务器响应所请求的内容,然后关闭连接。浏览器读取这些内容,并把它显示在屏幕上。

Web 服务和常规的文件检索服务(例如 FTP)有什么区别呢?主要的区别是 Web 内容可以用一种叫做 HTML(Hypertext Markup Language, 超文本标记语言)的语言来编写。一个 HTML 程序(页)包含指令(标记),它们告诉浏览器如何显示这页中的各种文本和图形对象。例如,代码

```
<b> Make me bold! </b>
```

告诉浏览器用粗体字类型输出 和 标记之间的文本。然而,HTML 真正的强大之处在于一个页面可以包含指针(超链接),这些指针可以指向存放在任何因特网主机上的内容。例如,一个格式如下的 HTML 行

```
<a href="http://www.cmu.edu/index.html">Carnegie Mellon</a>
```

告诉浏览器高亮显示文本对象“Carnegie Mellon”,并且创建一个超链接,它指向存放在 CMU Web 服务器上叫做 index.html 的 HTML 文件。如果用户单击了这个高亮文本对象,浏览器就会从 CMU 服务器中请求相应的 HTML 文件并显示它。

旁注 万维网的起源

万维网是 Tim Berners-Lee 发明的,他是一位在瑞典物理实验室 CERN(欧洲粒子物理研究所)工作的软件工程师。1989 年, Berners-Lee 写了一个内部备忘录,提出了一个分布式超文本系统,它能连接“用链接组成的笔记的网(web of notes with links)”。提出这个系统的目的是帮助 CERN 的科学家共享和管理信息。在接下来的两年多里, Berners-Lee 实现了第一个 Web 服务器和 Web 浏览器之后,在 CERN 内部以及其他一些网站中, Web 发展出了小规模拥护者。1993 年一个关键事件发生了, Marc Andreessen(他后来创建了 Netscape)和他在 NCSA 的同事发布了一种图形化的浏览器,叫做 MOSAIC,可以在三种主要的平台上所使用: Unix、Windows 和 Macintosh。在 MOSAIC 发布后,对 Web 的兴趣爆发了, Web 网站以每年 10 倍或更高的数量增长。到 2015 年,世界上已经有超过 975 000 000 个 Web 网站了(源自 Netcraft Web Survey)。