

11.5.2 Web 内容

对于 Web 客户端和服务端而言，内容是与一个 MIME(Multipurpose Internet Mail Extensions, 多用途的网际邮件扩充协议)类型相关的字节序列。图 11-23 展示了一些常用的 MIME 类型。

MIME 类型	描述
text/html	HTML 页面
text/plain	无格式文本
application/postscript	Postscript 文档
image/gif	GIF 格式编码的二进制图像
image/png	PNG 格式编码的二进制图像
image/jpeg	JPEG 格式编码的二进制图像

图 11-23 MIME 类型示例

Web 服务器以两种不同的方式向客户端提供内容：

- 取一个磁盘文件，并将它的内容返回给客户端。磁盘文件称为静态内容(static content)，而返回文件给客户端的过程称为服务静态内容(serving static content)。
- 运行一个可执行文件，并将它的输出返回给客户端。运行时可执行文件产生的输出称为动态内容(dynamic content)，而运行程序并返回它的输出到客户端的过程称为服务动态内容(serving dynamic content)。

每条由 Web 服务器返回的内容都是和它管理的某个文件相关联的。这些文件中的每一个都有一个唯一的名字，叫做 URL(Universal Resource Locator, 通用资源定位符)。例如，URL

`http://www.google.com:80/index.html`

表示因特网主机 `www.google.com` 上一个称为 `/index.html` 的 HTML 文件，它是由一个监听端口 80 的 Web 服务器管理的。端口号是可选的，默认为知名的 HTTP 端口 80。可执行文件的 URL 可以在文件名后包括程序参数。“?” 字符分隔文件名和参数，而且每个参数都用 “&” 字符分隔开。例如，URL

`http://bluefish.ics.cs.cmu.edu:8000/cgi-bin/adder?15000&213`

标识了一个叫做 `/cgi-bin/adder` 的可执行文件，会带两个参数字符串 15000 和 213 来调用它。在事务过程中，客户端和服务端使用的是 URL 的不同部分。例如，客户端使用前缀

`http://www.google.com:80`

来决定与哪类服务器联系，服务器在哪里，以及它监听的端口号是多少。服务器使用后缀 `/index.html`

来发现在它文件系统中的文件，并确定请求的是静态内容还是动态内容。

关于服务器如何解释一个 URL 的后缀，有几点需要理解：

- 确定一个 URL 指向的是静态内容还是动态内容没有标准的规则。每个服务器对它所管理的文件都有自己的规则。一种经典的(老式的)方法是，确定一组目录，例如 `cgi-bin`，所有的可执行性文件都必须存放这些目录中。
- 后缀中的最开始的那个 “/” 不表示 Linux 的根目录。相反，它表示的是被请求内容类型的主目录。例如，可以将一个服务器配置成这样：所有的静态内容存放在目录 `/usr/httpd/html` 下，而所有的动态内容都存放在目录 `/usr/httpd/cgi-bin` 下。