

HTTP详解1

Hello，大家好。

这里是**放弃编程**。

我们开始了一个全新的栏目叫做《HTTP详解》。

我们将要在这里面从最基础的讲起，逐步深入的了解网络传输。

这个栏目每次更新有可能有一篇文章，也有可能有两篇文章，如果有两篇文章，标有“进阶”字样的是适合有基础的小伙伴阅读，对于零基础的小伙伴可以选择只看没有“进阶”字样的文章。

我们今天要来了解一下网络是如何运作的，或者说现代浏览器（这里的现代浏览器指的是主流的浏览器，比如说Chrome、Firefox、Safari、Edge等，不包括IE）是如何工作的。

在说具体的运行机制之前呢，我们先来了解一下现代互联网的组成。

现代互联网的组成呢，大致上分为这几类：**终端、集线器、交换机、路由器、网络运营商、服务器**。

终端就是我们日常生活中经常使用的，比如说台式电脑，笔记本电脑，手机，Pad等等。

至于**集线器**和**交换机**则是一种传递网络的装置，我们将在之后详细的聊聊它们。

路由器，这个名字大家会很熟悉，在生活中使用Wi-Fi总不会少了它。但是，作为Wi-Fi装置只是它的一部分功能。

网络运营商我不说，大家也都明白是什么，生活中大家也都离不开这三家，也没少打交道。

最后这个**服务器**的概念就比较宽泛了，未来我们也会重点的聊一聊它的。目前大家只需要知道，所有你想访问的网页也好，图片也好，视频也好，都是存储在**服务器**中，而我们浏览网页的本质就是将存储在**服务器**上的网页下载到电脑上的浏览器中，然后浏览器负责将你下载的网页呈现出来。

所以我们浏览网页的过程就是：**终端 -> 集线器/交换机/路由器**（这里有可能不止一个） -> **网络运营商 -> 集线器/交换机/路由器**（这里有可能不止一个） -> **服务器**（这里也有可能不止一个，我们暂时默认只有一个） -> **集线器/交换机/路由器**（这里有可能不止一个） -> **网络运营商 -> 集线器/交换机/路由器**（这里有可能不止一个） -> **终端**。

如下图所示：

下面让我们用**终端**来打开一个网页。

首先打开一个浏览器，我们在浏览器的地址栏里面输入一个网

址：<http://www.fangqibiancheng.com>。（当然这个网址是我随便编的，大家就默认它是一个真实存在的网址就行了。）

输入完毕后我们按下回车键。

网页就开始加载了，然后直到加载完毕，网站在浏览器中显示在我们面前。

现在我们要来了解的就是从我们按下回车开始，整个网络是如何将你想要访问的网站，按照上面的访问顺序，最后呈现在你面前的。

我们先来看第一步。

我们按下回车键之后，浏览器就开始解析你刚刚输入的网址，即<http://www.fangqibiancheng.com>。

这个<http://www.fangqibiancheng.com>在计算机网络中有一个名词，叫做**URL**，翻译过来就是**统一资源定位符**。我们只用记住它叫做**URL**就可以了。

这里的这个**URL**是一个比较简单的**URL**，还有更复杂的**URL**，比如说<http://www.fangqibiancheng.com/home/index.html?search=a&t=1>

你看我们在后面又加上了许多的字母、符号。这些都叫做**URL**，简单来说，在浏览器的地址栏中能够识别的都叫做**URL**。

下面我们再介绍一个名词叫做**URI**，中文名称是**统一资源标识符**。你看**URI**和**URL**长得很像吧。像就对了，你可以理解为**URL**是**URI**的一种。

既然**URL**只是**URI**的一种，那么其他的**URI**都有什么呢？

这里我们来看一下官方定义的**绝对URI**的格式（对，你没看错，是官方定义的，由国际互联网资源管理的非盈利社团ICANN（Internet Corporation for Assigned Names and Numbers），中文名称叫做**互联网名称与数字地址分配机构**，下属的IANA（Internet Assigned Numbers Authority），中文名称叫做**互联网号码分配局**，来管理的。）：

```
1 http://username:password@www.example.com:80/dir/index.html?id=1#p1
```

很多小伙伴就疑惑了，上面这都是啥，为啥和我在浏览器里面看到的和输入的全都不一样呢？

我们从头开始看。

- **http:** - 这里的 **http:** 表示**HTTP**协议，目的就是告诉浏览器，访问 **http:** 后面的内容的时候应该使用**HTTP**协议。这里的协议大家可以理解为一种全世界通用的规则，这个规则中规定了应该如何告诉服务器想要执行的操作，就相当于写邮件时有一种固定的格式，当然邮件的格式是一种约定俗成，而HTTP协议则是白纸黑字记录下来的标准。当然了，除了HTTP协议之外，还存在其他的协议，比如说FTP协议。其实我们可以把这个部分理解为我们用户想要访问的功能，因为这里还支持邮件（mailto:）、新闻（news:）、电话（tel:）等等。
- **//** - 这里大家可以理解为分隔符，将协议与内容分隔开，方便识别和记录。
- **username:password** - 这里看英文名称大家也都能明白是用户名和密码，使用冒号分开。这个部分在使用**HTTP**协议中会被省略，省略时会后面紧跟着的 **@** 符号一起省略。因为不是重点，所以我们也不过多介绍。
- **www.example.com** - 这里用一个专有名词来称呼，叫做**域名**。也可以叫做**web服务器名**。web这个词本身就表示了“网络”的意思，所以我们在浏览器中看到的页面可以称作为**web页面**，对应的存储这些web页面的服务器就可以叫做**web服务器**，**web服务器名**就可以理解为web服务器的名字。因为不可能存在世界上所有的网页都放在一个服务器中的情况，所以对于访问不同的网站，这里的**域名**也是完全不一样的。我们通过这里来区分想要访问的网站是哪个网站。
- **80** - 这个部分叫做**端口号**，大家可以将**服务器**理解为一个现实生活中距离你很远的朋友的家，我们

要去朋友家拜访，就需要找到他家的地址，地址在这里可以看作是**域名**，而**端口号**就可以认为是朋友家的门牌号（单元号，楼层号，门牌号）。我们在浏览器地址栏中一般是看不到**端口号**这个部分的，原因是**端口号**这个部分是有默认值的，使用**HTTP**协议就是**80**，使用邮件就是**25**，使用**HTTPS**就是**443**之类的，这就好比你家坐落在一个很大的庄园内，客人来你家拜访，只要给客人地址就可以了，并不需要告诉客人从前门进还是从偏门走，客人来拜访时自然而然的第一选择就是正门。**端口号**和**域名**之间使用冒号（:）分割，省略**端口号**使用默认**端口号**时将前面的冒号一并省略。

- **/dir/index.html** - 着表示你要访问的内容在服务器中的地址，就好比，你去别人家借扳手，你就需要告诉邻居你是来借扳手的，放到这里也就是告诉服务器，你想看服务器中的哪些内容。其中 **/dir** 部分我们叫做**目录**，这个目录可以很长，也可以很短，长了可以叠加 **/dir/dir/dir**，最短就是 **/**。后面的 **/index.html** 就是**文件**，即我们要访问的文件。这里有一个概念，就是所有你可以访问到的、网络上的内容几乎都可以当作不同的文件来处理，比如说一个网页就是一个 **html** 文件，一首歌就是 **mp3** 文件，一个视频就是 **mp4** 文件之类的，一个网站我们可以看作是不同文件的集合。**/dir/index.html** 这里表示的就是我们要访问 **index.html** 这个文件。当然了，**文件**部分可以忽略，因为在服务器上通常会设置一个默认的文件，当你没有指定**文件**时会自动访问前面提供的**目录**下的默认文件。
- **?id=1** - **查询字符串**——所有我们见到的显示出来的文字都叫做字符串——所以这里表示查询用的文字。查询什么呢？拿上面借扳手来举例，扳手也有很多种类，有大的有小的，有梅花扳手，有活扳手，有两用扳手，这里的**查询字符串**就是告诉邻居你要借的扳手是什么样的扳手，这样邻居就会根据你的需求借给你对应的扳手。当然了，也会出现你想借扳手，邻居并没有的问题，这个情况我们将在之后讲解会出现什么样的处理。
- **#p1** - **片段标识符**。这一部分其实并不归网络处理，而是在浏览器内部就给处理了，所以这里我们并不作详细说明。

上面我们对**绝对URI**的各个部分做了说明。

当浏览器拿到 **http:** 开头的符合规范的字符串就会进行下面一步——根据**HTTP**协议拼接HTTP请求的消息。

下面我们会对**HTTP**协议，以及什么叫做HTTP请求、如何拼接消息做进一步说明。