环球网WWW及其使用

孙淑玲 编著

中国科学技术大学出版社

环球网 WWW 及其使用

孙淑玲　编著

中国科学技术大学出版社

1996 · 合肥

内 容 简 介

WWW(World Wide Web)是目前最流行的新型查询工具，使人们能在全球范围内共享 Internet 上的资源和知识。用户可以在 Internet 上查找各 WWW 站点的超文本、超媒体信息，并且无痕迹地把它们集成在一起。本书在简要地介绍 WWW 的发展情况和基本概念后，系统地介绍了 WWW 中的主要协议和标准：超文本传输协议 HTTP，超文本标记语言 HTML，统一资源地址 URL，公共网关界面 CGI，其中包括较多典型的示例。此外，还叙述了建立 WWW 站点应该考虑的问题以及服务器 CERN-http 和浏览器 Netscape 安装事宜。最后提供一批关于 WWW 和相关资源的 URL 地址。

本书适合于广大计算机爱好者，大专院校师生，工程技术人员和科研工作者。

图书在版编目

JSH16/13

环球网 WWW 及其使用/孙淑玲 编著 - 合肥：中国科学技术大学出版社，1996
ISBN 7-312-00772-4

I 环球网 WWW...
II 孙淑玲 编著
III ①环球网 ②超文本传输协议(HTTP) ③公共网关界面(CGI)
IV TP

*

中国科学技术大学出版社出版发行
(安徽省合肥市金寨路 96 号，邮编：230026)
中国科学技术大学印刷厂印刷
全国新华书店经销

*

开本：787 × 1092/16　　印张：13　字数：316 千
1996 年 8 月第 1 版　　1996 年 8 月第 1 次印刷
印数：1 — 6 000 册
定价：14.50 元
ISBN 7-312-00772-4/TP · 129

前　言

计算机从诞生至今只有短短50年历史，它经历了大型机垄断、个人计算机革命两次浪潮的洗礼，今天又迎来了第三次浪潮。这次浪潮的特点是以强大的计算机网络促进计算技术的发展。个人计算机将从网络获取大量信息，更多地依赖从网络获得的“集体智慧”。也就是说可以与任何国家，任何地方的人直接沟通，实现全球范围内的知识共享。通过网络互相传递声音、数据、图像或影视多种媒体信息，从而使百万大脑产生的知识和信息创造性地联系起来。

国际互连网络 Internet 是实现上述目标的基础。在 Internet 上用网络中心高性能服务器支持为数众多的个人计算机，把世界连成一个整体。 WWW 是在 Internet 上广泛使用的一个超文本信息和通信系统。在 WWW 中以客户机-服务器模式进行数据通信。浏览器作为 WWW 客户机可以使用多种协议并且访问多媒体信息，能依据用户的需要组织和传递信息，从而 WWW 成为较理想的知识共享手段。

江泽民同志指出：“四个现代化，哪个也离不开信息化”。信息化是实现我国四个现代化的关键。目前，全世界所有国家都站在下一世纪的信息公路的入口处。如何把握这一机遇，迎头赶上是落在我们肩上的历史性任务。在我国分组数据交换主干网建立之后，各部门、各企事业单位都纷纷建网。 WWW 系统能给予信息提供者以发布信息的能力和在全世界范围内及时收集信息的能力。从用户来讲， WWW 是一个动态视窗，了解世界各地的人和组织在想什么做什么。

本书的目的是把 WWW 系统介绍给广大读者。书中全面地叙述了 WWW 的性质和特点， HTTP 协议， HTML 语言， CGI 标准，并配有丰富的例子和信息资源。它适合多种层次，多种目的的读者。希望能促进我国网络的普及和应用。网络和通信是我们研究所的主要研究方向。目前已经开展了 WWW 系统和 Java 语言方面的研究工作。推广使用新的通信系统也是我们义不容辞的责任。

本书的第三章和第四章由周虹同志完成，刘祥春和朱雪松同志提供并测试了第七章的全部例子和第八章部分内容。在写书过程中与陈意云、韩梅、陈凌宇同志进行了有益的讨论。我在此一并深表谢意。

孙 淑 玲
1996.2.16

作者单位：中国科学技术大学计算机网络与系统研究所
通讯地址：安徽合肥中国科学技术大学计算机系
邮政编码：230027
E-mail: slsun@dawn1.cs.ustc.ac.cn

目　次

第一章 引 言

当前，一个兴建“信息高速公路”的热潮正在席卷美国、日本和西欧等发达国家。人们都在憧憬这一美好的未来，许多科学家和工程师正在为之努力奋斗。目前已经建立的国际计算机互连网络 Internet 联系着全世界 137 个国家和地区，拥有 2200 万以上的用户。Internet 已成为发展迅猛的电子领域，开拓者蜂拥而至。每天 Internet 的规模、范围及重要意义都在增长。每天人们都给网络增加更多的功能，更大的安全性和能力。Internet 是未来信息高速公路的原形与基础。 1995 年在美国拉斯维加斯召开了全球性计算机业盛会 Comdex/fall'95 。 IBM 公司总裁兼首席执行官 Louis V.Gerstner 在开幕式上所做的长篇演说中指出: “我们正站在计算的下一个重要时期的门槛上。计算机将进入一个以网络为中心的计算的新时期”。在会上特别开设了 Internet 专题研讨会，其中对于开发实用 WWW 战略的意义做了深入的讨论。

从网络通信技术角度看， Internet 是一个以 TCP/IP 协议连结各国家、各部门、各机构计算机网络的数据通信网。从信息资源角度看， Internet 是一个把各个领域各种信息资源连结为一体的数据资源网。

Internet 提供了三项基本服务: 电子邮件(E-mail)、远程登录(Remote Login)和文件传送(File Transfer)。在此基础上，为了帮助用户从浩翰的信息海洋中容易找到自己需要的信息，相继开发了几个查询工具。它们是:

(1) Archie 世界上有 1500 多个宿主机装有 Archie 系统。 Archie 管理着一个文件名数据库，记录着每个文件的名字和存放地址。该数据库分布在世界不同地区的 40 多个 Archie 服务器上。这些服务器每天对自己管辖区域内的计算机文件搜索一遍，以便及时核实、修改、补充数据库的内容。这 40 多个 Archie 服务器每月相互交换一次信息，以保证每个 Archie 服务器都有一个经过更改的、完整的、一致的文件名数据库。

(2) Gopher Internet 上大约有 5000 个 Gopher 服务器。它也是一个文件定位应用软件。用户要用一些查询命令逐步缩小查找范围。 Gopher 是由具有层次结构的菜单驱动的。首先出现的是主菜单，它给出很粗的分类表。用户只需按“↑”、“↓”、空格键或数字键，在菜单的引导下迅速靠近用户的目标。找到文件地址后就可以把它取回来。如果是文本型文件，还可以当场打开看看是否是所要的文件。

(3) WAIS 它是广域信息服务器。 WAIS 查找的是各种文件内部的信息。在各处的 WAIS 站点管理人员建立描述文件内容的索引。被索引的文件信息收集到相关的数据库中。用户选定要查找的数据库后，输入查询命令就会得到一批满足条件的文件。

(4) WWW WWW 是环球网。它是基于超文本的信息查询和信息发布工具。特色在于为用户提供一种友好的信息查询界面。用户仅仅需要提出自己的查询要求，具体到什么地方查找，怎样取回都由 WWW 自动完成。用户需要做的事只是用鼠标器点击显示屏上高亮度或

有下划线的词语，就把与该词语相关联的文件取回并且显示在屏幕上。用户不必关心这些文件存放在 Internet 上的哪台计算机上。通过一级级跳越式查询，用户就方便地找到自己希望得到的信息。当然用户也可以返回到原来某个阅读位置，然后再顺序向下阅读。在 WWW 中，除了可以得到文本信息外还可以得到图像、声音和影视信息文件。另外，利用 WWW 能够方便地在 FTP 、 E-mail 、 Gopher 、 WAIS 等多个系统之间切换。用户通过自己的起始页(Home Page)提供自己的信息，让 WWW 上其它读者共享。

本书的目的是介绍 WWW 的实现机制和使用方法。

1.1 WWW的由来及现状

WWW 开始于欧洲粒子物理实验室(CERN)。该实验室是欧洲 12 国联盟高能物理学家的一个组织。 1989 年， CERN 的物理学家 Tim Berners 提出应该建立一个有效的分布式信息系统，使得高能物理协会的科学家们用它来传递新的思想和新的研究成果。他明确指出，该系统应当使用一种全新的方式——超文本(Hypertext)在计算机网络上传送文件。这里链不仅用来表示相互关联的信息而且还可以用来提交信息。 1990 年，第一个 WWW 软件在 Steven Job 的 NeXT 计算机系统上诞生。该软件能够在 Internet 上阅览和传送超文本文件。在随后的几年中，这个系统又有较大扩展。令人们难以相信的是 WWW 以每年 30 倍的速度在增长。在 1993 年世界上只有几百个 WWW 服务器而 1994 年就有 1 万个。据保守的估计，每天新建的 WWW 服务器达 500 个。其中教育与商业部门分别占 50%和 25%。 WWW 系统发展如此之快超出了人们的预料。从 1993 年到 1995 年已经召开了 4 次 WWW 国际会议。某些专家预测，到本世纪末 Internet 和其核心部分 WWW 将拥有 10 亿用户。 Internet 的三位创始之一 Tim Berners-Lee 说：“ Internet 、特别是 WWW 正从出现时的一种纯应用发展为我们进行通信、学习、计算和商业的潜在的信息空间。许多公司参加 W30 是因为他们要求那种空间稳定可靠，并有所发展。他们认识到， WWW 作为一条高速公路和一个市场，必定会成为使其产品有竞争力的开路先锋。”

WWW 快速增长的原因在于:

(1) Internet 发展迅速。现在 Internet 名符其实地成为全球通信网络。客观上需要一种能链接网上所有资源的应用系统。

(2) 交互性超文本、超媒体和多媒体软件的开发，使得不掌握多媒体技术的人也能在不同的文件和媒体资源之间建立连接。另外，开发了一批出色的 WWW 浏览器（如 Mosaic， Netscape 等），使得用户只简单地按几下键就能在 Internet 上漫游。

(3) 多媒体技术已经走进办公室进入家庭。商业部门很快接受了多媒体概念，用它培训雇员或向客户发布电子文件。大公司用他们自己的计算机网络向世界各地的雇员发送电子新闻、电子电话簿和产品信息。各个部门、科室、工作小组之间共享数据、情况简报和某项工程的细节资料等。用超文本软件都能把它们链接在一起。

(4) Internet 把众多计算机网络连接在一起，其间那一个都不是 Internet 的拥有者或监控机构。它们处于一种无政府状态。在这一点上， WWW 与 Internet 十分相似。任何个人、公司、组织都不是 WWW 的主人。 WWW 是拥有数亿用户的分布式系统。每个用户都能

撰写超媒体文件，成为向世界提供信息的主体。这些用户可能是学者、大中学生、市场专家、律师、音乐家、园艺师等各行各业的普通人。

多媒体信息进入巨大的计算机网络中，需要有统一的方式去访问包含着各种类型知识和信息的数据库。而这些数据库应用的范围极其广泛，例如：娱乐、教育、商业、通信等。WWW 就是把它们都能链接在一起的大拼盘。

美国斯坦福研究所(SRI)的学者对 Internet 和 WWW 的使用情况做过一次统计。他们详细比较了政府部门、研究单位、教育部门、团体组织网点的使用情况。结果如下：

	使用 WWW	使用 Internet
美国教育部门	49%	27%
美国商业部门	20%	26%
美国政府部门	9%	6%
其它国家与领域	20%	41%

在 1300 份问卷调查中，使用 WWW 的用户情况如下：

使用WWW的用户	占的百分比
20—30岁之间的人	56%
男人	94%
北美地区	69%
在职人员	45%
研究生	22%

我国在世界银行贷款项目 中国国家计算和网络实验室 NCFC 的基础上建成了 ChinaNet 。它用一条 64Kbps 专线，经由 Sprint 国际路由器与 Internet 主体 NSFNET 相连。ChinaNet 是全球互连网络的中国部分。它由国内众多网络互连而成，主要包括：

网络名称	英文名称
中国科学院网	ASNET
北京大学校园网	PUnet
清华大学校园网	TUnet
中国科技大学校园网	USTCnet
高等学校网络	Canet
国家科委网	SSTCnet
生态系统研究网	CERNet
国家海洋环境预测研究网	MEFnet
北京科技信息协会网	BSTIS
国家洪涝灾害控制网	NFCWAN

此外还有科学院高能物理所、微生物所、上海地区、武汉地区、澳门地区网等。在科学院高能物理研究所的 Home Page 中列出我国目前拥有的 WWW 服务器一览表，地址为：

http://www.ihep.ac.cn/china-www.html

http://www.w3.org/hypertext/DataSource/www/servers.html

在清华 WWW 服务器中有北京电子杂志“神州学人”，地址是 http://www.chisa.edu.cn 。1995 年全国各地进入 Internet 的用户已上万。目前与 Internet 联网的电子期刊已有 7 — 8 家国内报刊。在试用期间，只要进入“ http://www.em.co.cn ”即可免费浏览这几家电子报刊。预计 96 年我国进入 Internet 的用户将突破 10 万大关，信息服务业的 WWW 服务器及各种中文信息资源也将相继在 Internet 上亮相。

1.2 WWW系统的特点

现代通信技术把全球都联系在一起。WWW 是人们在世界范围内查找信息和共享知识的理想工具。WWW 系统的特点是：

1. WWW 是基于 Internet 的漫游系统

WWW 连接 Internet 上的许多资源。从使用的角度看，它比 Archie 、Gopher 、WAIS ，更能激发人们在 Internet 上漫游的欲望。WWW 把各种形式的信息，如文本、图像、声音、影视无痕迹地集成在一起。在用户使用 WWW 时，可以高效率地在数千台计算机之间，在各种系统应用程序(FTP ， Telnet)之间，在各种信息形式之间自由跳跃。WWW 系统摒弃了费解的计算机命令，代之以按动鼠标器，沿着超文本链实现上述跳跃。最直接的好处是容易在大型文件中旅行。例如：一个公司提供给用户的某种远程通信交换机设备手册估计多达 10 万页。如果做成 WWW 形式的电子文件，在文件内部设立众多热点主题，那么用户可点击一个主题，在链的引导下到达该主题所在的段落。又如：在 Internet 上有全文存放从狄更斯到马克吐温时代大约 100 本经典文学名著的数据库。用户使用 WWW 系统可以在自己的计算机上阅读，也可以取回一部分放在自己机器上待闲暇时再阅读。实际上，用 WWW 漫游还可以找到许多电子期刊和杂志。只要几分钟一个文件就能从一个国家“飞”到另一个国家。人们将情不自尽地发出感叹：“世界原来这么小!”。又例如：英国新闻服务公司建立了自己的 WWW 服务器。它把从 BBC 得到的新闻按超文本形式编辑在一起。过去花半小时找到消息,而阅读只需一分钟。现在总共只要花几分钟就行了。

2. WWW 是信息分布式系统

在 Internet 上有 1 万多个 WWW 服务器，有几亿份可供使用的文件。这些服务器和文件分布在世界不同的地方。文件中的超文本链是一种电子指针，它连接到其它 WWW 服务器上的信息或资源上。当用户用鼠标器点击超文本链的时候，用户并不知道该文件放置在哪台计算机上。从而使得用户就像在电视机上更换频道一样，容易地在世界 80 多个国家里的文件和资源间任意走动。值得一提的是，这种电子指针可以指向若干非介质的信息上。例如：一个超文本链可以打开一个 FTP 服务器或用 Gopher 系统进行一次信息查询。

3. WWW 是一个交互式的超媒体系统

WWW 为用户提供基于超媒体的信息和文件。WWW 资源包括文本、图像、声音、影视，

等各类信息，使用起来十分方便。超媒体信息使得文件具有活力，也使得计算机成为一个多媒体设施。这样的计算机比起收音机、电视机更具魅力。例如：一个电子自动修理手册中有一节描写怎样调试碳棒。用户使用鼠标器点击屏幕上的一个图标就会放一段录象，展示整个调试过程。接下去是一段文字说明“调到声音变成圆滑为止”。可能用户不知道哪种声音才算圆滑，那么用户点击另一个图标就会播放“圆滑声音”录音带。不难想象这样做用户一定学得快、记得牢、收获大。

WWW 与电视机是不同的。电视节目由制作人事先确定好，用户只能按步就班地得到信息。在使用 WWW 时，用户自己主动控制着要看的信息。在超媒体的引导下，用户本身成为开拓者。自己判断此刻哪个信息最重要，应该深入探讨哪个问题等等。

WWW 支持多媒体信息是指以下三种含义：

- 文本中的信息类型可以是不同的，也可以在一种媒体信息中包含另一种媒体信息。
- 允许用户从远程计算机上取回各种类型文件，在自己的计算机屏幕上阅读或播放。
- 所有类型信息(包括图像或图像的某几个部分)可以链接到其它的信息片段上。

1.3 WWW的未来

早期，CERN 在定义 WWW 中起了关键作用。现在 CERN 与 MIT 计算机科学实验室联合建立了一个 WWW 组织，称为 W30(即 World Wide Web Organization)。它的宗旨是实现 WWW 最终目标，即使得用户使用任何计算机都可以用 WWW 访问其上的资源。MIT 把介质实验室、人工智能实验室等几个研究组都投入到 WWW 研究中来。以这两个单位为中心联系周围其它的研究所和 WWW 站点，围绕着上述目标开展工作。

W30 在开始阶段主要是研制新的国际数据实体和修正 WWW 标准。希望在查找信息时更加容易。由于 WWW 正在变成一个巨大的电子图书馆，采用标准化方法对信息进行分类必将大大改善 WWW 信息的可访问性。从现在的 WWW 标准出发，新的标准应该保证众多公司开发的软件是兼容的。

为了 WWW 的全球商业应用，W30 将规定有关安全性、私有性及资金电子转账等各种条款。该项目的资金来自美国政府、欧洲联盟和国际上一些大公司。虽然 W03 本身不是标准化组织，但是它的开发和研究工作将会引发出许多新概念和新技术值得大家关注。它们主要研究的课题有：

- 命令和语法与语义的描述；
- 网络传输协议HTTP；
- 超文本与超媒体所用的数据格式HTML；
- 用于压缩和安全的编码技术；
- 用于填表的协议以及用于传输合法连接文件的协议；
- 增强WWW功能（PROXY服务器，CACHING快速缓存，复制与优化请求路由）；
- 使用另一种高速网络技术ATM。

WWW 现在已经十分普及，当前仍然保持着高速度增长的势头。它的未来到底会发生什么尚有较大的不确定性。从近期来看，可以研制适用于多种平台的 WWW 浏览器和 WWW 集成工具。对人们已经熟悉的文字处理器和桌面印刷系统，浏览器应该提供转换工具，自动地

生成 HTML 格式文本。 SUN 公司在推出 HotJava WWW 浏览器后，又不失时机地与 Netscape 公司于 1995 年 12 月共同推出一种开放的，跨平台的对象描述语言 JavaScript 。它所产生的联机应用软件能够把客户机和服务器上的对象和资源联系在一起。 HTML 文件作者和企业软件开发人员能用 JavaScript 描述客户机或服务器上运行的各种对象的行为。由 JavaScript 形成的联机应用软件可以通过 Internet 动态地表达信息并与用户进行交互。该语言已提交 W30 委员会审查，以便成为描述语言标准。人们相信有了 JavaScript 语言后将使 WWW 的应用上升到新的高度。

人们有了浏览器和与之相应的指导手册以后，如何通过 WWW 得到用户想要的所有关于某个课题信息，仍然是一个尚须认真解决的问题。美国 MIT 和华盛顿大学的研究人员做了一些探索性的工作，称之为电子机器人或知识人。知识人是一种为用户进行 WWW 查询的程序。用户准确地告诉知识人要查找什么，知识人就能设计出经由网络到达相应主机的方案，并顺利取回信息或资源。这种知识人可用于建立商业上的“比较价格购物”系统。知识人首先查询销售某特定产品的零售商，保存关于该产品的详细信息(特别是价格)，形成报告以便帮助用户做决断. 当然开发知识人的公司还有一些技术问题要解决，例如知识人可能传播病毒，威胁对方计算机的安全。另一方面，假设网络上有几百万个知识人在行动，显然网络上信息传送速度将会急剧下降。如何解决堵塞问题就显得十分重要。现在已经有两个知识人在 WWW 上运行

- MIT Matthew Grey 研制的WWWWander。它通过WWW的链找到相应WWW站点，提供可用的超文本文件的信息。
- WebCrawler是华盛顿大学Brian Pinkerton研制的. 它主要收集在WWW服务器上驻存的特定文件信息，并为这些文件建立索引 用户使用关键字查询该索引。

第二章 基本概念和协议

WWW 以客户—服务器模式为基础，多数客户用图形界面浏览器向服务器提出请求。如果读者是第一次接触 WWW 和超媒体，特别是从来没有使用 UNIX 的经验，那么读者需要先了解一些专业词汇:

(1) 对象(object) WWW 中一个对象是指文件(file)为形式的数据块。文件可以包含任何类型的数据。例如: 文本、图像、影视、声音以及它们的组合。文件也可以是可执行程序。通常用户向 WWW 请求数据时，取回的文件对象(document object)包含文本、在线图像和指到其它对象的链。

(2) 页(page) 页是一个连续的数据片。它像一个字处理文件。用户可以用鼠标器拉动滚条去看当前屏幕上看不到的内容。

(3) WWW 站点(WWW site) 站点是通过链连接在一起的在概念上相关的一组页面。

(4) 浏览器(browser) 要取回 WWW 文件或其它对象，就需要有个浏览器。浏览器是一个软件程序，它知道怎样在 WWW 系统中沿着链漫游，怎样与 WWW 服务器通信，怎样取回数据。浏览器也称为客户。

(5) 服务器(server) 服务器也是一个软件程序。它管理 WWW 站点上的数据，回答浏览器的请求。

(6) 用户(user) 一般用户是使用计算机存取信息的人。然而用户不一定是实实在在的人。用计算机的术语来讲，用户就是计算机上的一个户头。户头应该有一个名字和关于存取文件权利或运行程序权利等方面的信息。

2.1 客户－服务器模式

WWW 是建立在客户—服务器模式之上的。客户和服务器是相互通信的一对程序。客户连接到服务器上并请求一段信息。WWW 浏览器是连接到 WWW 服务器的客户。

服务器的任务是等待外面客户来连接，听取客户的请求并为这些请求服务，对客户给出回答。为了使客户与服务器相互理解，它们双方必需遵守协议(protocol)。如果客户没有按要求的方式提出请求，则服务器不能给出适当的回答。类似地，如果服务器不根据协议去回答，那么客户也无法理解。当然，客户和服务器要遵守同一个协议，使得双方具有共同的准绳。一般人们都期望协议高效、健壮、功能强大。如果客户与服务器不严格遵守协议，那么这些优点都白费了。

在 WWW 出现之前，Internet 上已经有了许多协议，如 FTP 、Archie 、Gopher 等。它们每个都定义了从一台机器经由 Internet 到另一台机器传输信息以及查找和取回信息的方法。这些协议对 WWW 设计起了很大作用。在 WWW 系统中，仍然能够使用这协议。所

以，我们说 WWW 是建立在这些协议之上的。更为重要的是 WWW 把用户与协议分隔开来。从外观上看，用户面对的是统一界面，用户察觉不出信息是按那个协议取回来的。此外，WWW 不限制传送对象的类型。

WWW 主要用以下协议和标准定义:

- URL——统一资源地址
- HTTP——超文本传输协议
- HTML——超文本标记语言
- CGI——公共网关界面

WWW 中使用这协议和标准进行信息定位、信息存取和信息显示。用在公共登录格式(Common log Format)的标准也应该遵守。人们根据它编写服务器登录文件的分析程序。

2.2 URL

URL (Uniform Resource Location) 是一种标准化的命名方法。经由各种不同的协议，对 Internet 上任何地方的信息都可以用 URL 定位或取回。 URL 可以指定 FTP 文件传输，寻找新闻信息，定义用户的 E-mail 地址。标识 HTTP 文件和其它类型数据。在第六章将给出详细描述。

2.3 HTTP

HTTP 是 WWW 上用于发布信息的主要协议。它既简单又有高度灵活性. 为了从服务器把用户需要的信息发送回来， HTTP 定义了简单事务处理。简单事物处理由以下四步组成:

- 客户与服务器建立连接;
- 客户向服务器递交请求。在请求中指明所要求的特定文件;
- 如果请求被接纳,那么服务器送回一个应答。应答中至少包括状态编码和该文件内容;
- 客户或服务器断开连接。

HTTP 的主要目的之一是提供一个简单算法，使得服务器能迅速地为客户做出应答。为此， HTTP 应该是无状态协议，即从一个请求到另一个请求不保留任何有关连接的信息。 FTP 协议在这点上与 HTTP 协议截然不同。在 FTP 中是保留状态的。用户使用 FTP 时，用户能改变工作目录。 FTP 服务器记下原来目录后，又去满足用户的下个请求。而在 HTTP 中不能记住原来目录。实际上，人们可以在 HTTP 之外保存状态。许多 WWW 站点中状态编码就是保留状态的一种方法。此外，还可以用 CGI 程序修改或存放用户状态文件或数据库。

每次连接 HTTP 只完成一个请求。这点也与 FTP 协议不同。在一次请求完成之后，服务器与客户间的连接断开。用户再想取另一个文件还要重新与服务器建立连接。例如: 户要加载有多个在线图像的 HTML 页面。对于每个图像都要单独构成一次连接。虽然一次连接的时间开销并不大，但是对于远程站点或负载较重的站点来说就会花费较多时间，甚至

影响连接的实现。

有些浏览器只有在一次连接完成以后才能开始另一次连接。然而现在许多新的浏览器(如 Netscape)能并行地打开多个连接，同时取回多个文件，大大提高了工作效率。然而这会带来新的瓶颈问题。

在 HTTP 0.9 和 HTTP 1.0 之间主要的区别是在灵活性上。 HTTP 1.0 在以下两方面有较大改进:

1. 加入事务处理头

所谓头就是在主要数据前加入一个信息块。它是关于要检索的信息（或要传送的信息）的信息，称为元信息。在所有的事务处理中，不管是客户的请求还是服务器的应答统统加上头信息。当用户向服务器请求某数据时，头信息包括: 提出此请求的浏览器名，浏览器能接受哪种类型数据，浏览器懂得哪种语言等; 当服务器回答一个请求时，伴随着所取的数据还应该返回一个回答头。它包括: 该请求完成的状况，返回的信息长度，信息的类型，内容用哪种语言书写的，这段内容最后一次修改的日期等。例如:

```
HTTP/1.0 200 Document follows
Mine-Version:1.0
Server: CERN/3.0 pre 6
Data: Monday , 06-Mar-95 21:46:05 GMT
Content-Type: text/html
Content-Longth: 2848
Last-Modified: Thursday , 02-Mar-95 23:05:33 GMT
```

2. 增加新的方法

当客户向服务器提出一个请求时，要指定一种请求信息的方法。当然使用哪一种方法跟它的目的有关，也与服务器能力有关。在 HTTP 0.9 中只能使用 GET 方法。在 HTTP 1.0 中又加入了 POST 、 HEAD 、 PUT 、 DELETE 、 LINK 、 UNLINK 方法，共有七种方法。

2.4 HTML

HTML (Hypertext Markup Language)是一个按 SGML (Standad Genaral Markup Language) 定义的语言。它和其它标记语言（如 Latex ）一样，采用做记号的方法定义一块文本的特殊格式。在 HTML 中，标记是对文本中的一段进行语义标记。它不是具体呈现在浏览器上的物理标记。物理上如何实现留给浏览器自己决定。例如:

```
<H1>  This is the first level title  </H1>
```

表示的是由<H1>和</H1>括起来的一段文字用一级标题形式显示。这里每个字符的大小、字体、在屏幕上摆放的位置都由浏览器自己确定。所以，同一个 HTML 文件在不同浏览器中呈现的外观形式可能不同。

除语义标记外， HTML 提供了一个文件（对象）与另一个文件（对象）以及在一个文

件内的不同位置之间建立超文本链接功能。使用超文本链，允许用户任意地从一个话题转到另一个话题。正是由于超文本链构成了 WWW 系统中信息联系网。

HTML 1.0 是最早的 HTML 语言版本，它包括链和一些简单的标记。 2.0 版本能支持 FORM， 3.0 版本进一步支持 TABLE，数学公式和一些控制物理显示的功能。 Netscape 在 3.0 版本基础上又扩充了一些控制呈现方式的标记，如文本居中，字体大小等。

2.5 CGI

WWW 上链接着数目庞大品种繁多的信息，倍受人们青睐。然而 WWW 的多数信息是静止的。文件中的信息只是在站点管理人员干予或修改后才能发生变化。这种静态页面缺少交互性。能取到什么信息完全由信息提供者决定，用户处于被动的地位。

现在许多 WWW 站点已经能够让它的访问者达到半主动水平。人们也能把 WWW 看成交互性的媒体。从在线购物到远程机器人都充分体现了 WWW 交互性。只有靠 WWW 服务器的交互功能才能吸引用户，提高他们参与的积极性。 WWW 交互功能是靠 CGI 程序实现的。 CGI 程序使得信息网关、反馈机制、查询数据库、在线购物以及个人化文件成为可能。

CGI （ Common Gateway Interface ）是把客户程序与服务器程序结成一体的一个标准。最初，每个服务器都有自己的标准。这样，为一个服务器写的程序要想在其它服务器上使用必需做较大的修改。要是能形成一个公共标准，为一个服务器写的程序就能在任何服务器上运行。 CGI 就是完成此项任务的一个标准。 CGI 定义了一个界面。通过这个界面，服务器可以向 CGI 程序送信息， CGI 程序也可以向服务器回送信息。 CGI 是在 WWW 上建立交互功能的主要工具。

2.6 公共登录格式

公共登录格式（ Commom Log Format ）是访问登录文件的标准。有了它就能为任何服务器编写分析登录文件的程序。

公共登录文件规定了两件事:

(1) 登录什么信息及信息出现的顺序;

(2) 为了识别出这信息，应引入哪些分界符。

服务器程序自动地把有用信息放入登录文件中。这信息主要是提出请求的机器名、日期、时间以及请求是否成功等等。若对服务器软件稍加修改，登录文件还可以包括提出请求的用户名。下面是一个登录项的例子:

```
loki.netgen.com--[03/Mar/1995:17:48:12 f 0500] "GET http://www.
netgen.com/ HTTP/1.0" 200 2848
```

其中各部分含义是:

* loki.netgen.com是发出请求的客户所在的主机名,或者是远程客户的DNS或IP地址。
* 第一个 "-" 位置应该是Indentd 返回的信息。这里用 "-" 表示没有信息返回。
* 第二个 "-" 位置应该是UserName。为了安全进行认证时用户要送来UserID。这里