第一步是实现处理HTTP/1.0 GET请求的基本顺序代理。其他请求类型(如POST)是严格可选的。

*启动时，您的代理应该侦听一个端口上的传入连接，该端口的编号将在命令行上指定。一旦建立了连接，您的代理应该从客户机读取请求的全部内容并解析请求。它应该确定客户端是否发送了有效的HTTP请求;如果是这样，它可以建立自己到适当web服务器的连接，然后请求客户机指定的对象。最后，您的代理应该读取服务器的响应并将其转发给客户机。*

**4.1 HTTP/1.0 GET请求**

当终端用户在web浏览器的地址栏中输入一个URL(例如http://www.cmu.edu/hub/index.html)时，浏览器将向代理发送一个HTTP请求，该请求以可能类似于下面的一行开头:

GET http://www.cmu.edu/hub/index.html HTTP / 1.1

在这种情况下，代理应该将请求解析为至少如下字段:主机名www.cmu.edu;

以及路径或查询及其后面的所有内容，/hub/index.html。通过这种方式，代理可以确定它应该打开到www.cmu.edu的连接，并发送自己的HTTP请求，以以下格式的一行开头:

HTTP / 1.0 /hub/ index . html

*注意，HTTP请求中的所有行都以回车符' \r '结束，后跟换行符' \n '。同样重要的是，每个HTTP请求都以一个空行结束:“\r\n”。*

在上面的例子中，您应该注意到web浏览器的请求行以HTTP/1.1结束，而代理的请求行以HTTP/1.0结束。现代的web浏览器会生成HTTP/1.1请求，但你的代理应该处理它们，并将它们作为HTTP/1.0请求转发。

*重要的是要考虑到HTTP请求，即使只是HTTP/1.0 GET请求的子集，也可能非常复杂。教科书描述了HTTP事务的某些细节，但是你应该参考RFC 1945来获得完整的HTTP/1.0规范。理想情况下，根据RFC 1945的相关章节，您的HTTP请求解析器将是完全健壮的，除了一个细节:尽管规范允许多行请求字段，但您的代理不需要正确地处理它们。当然，您的代理永远不应该因为格式错误的请求而过早地中止。*

**4.2请求头**

请求头是HTTP请求中非常重要的元素。报头本质上是在HTTP请求的第一个请求行之后逐行提供的键值对。对于这个实验来说，特别重要的是主机、用户代理、连接和代理连接头。你的代理必须对列出的HTTP请求头执行以下操作:

*•总是发送一个Host报头。虽然HTTP/1.0规范在技术上不允许这种行为，但是有必要从某些web服务器(特别是那些使用虚拟主机的服务器)中诱导出合理的响应。*

*主机头描述了您的代理试图访问的web服务器的主机名。例如，要访问http://www.cmu.edu/hub/index.html，你的代理会发送以下头信息:*

*主持人:www.cmu.edu*

*web浏览器可能会将自己的主机头附加到HTTP请求上。如果是这种情况，您的代理应该使用与浏览器相同的主机头。*

*•总是发送以下User-Agent报头:*

*用户代理:Mozilla / 5.0 (X11;Linux x86\_64;房车:10.0.3)壁虎/ 20120305 Firefox / 10.0.3*

*报头提供在两个单独的行，因为它不适合作为一个单行在writeup，但你的代理应该发送报头作为一个单行。User-Agent头标识客户机(根据操作系统和浏览器等参数)，web服务器通常使用标识信息来操作它们提供的内容。发送这个特定的User-Agent: string可以在内容和多样性方面改进您在简单的telnet式测试中得到的材料。*

*•总是发送以下连接头:*

*连接:关闭*

*•总是发送以下代理连接头:*

*Proxy-Connection:关闭*

Connection和Proxy-Connection头用于指定在第一个请求/响应交换完成后连接是否保持活动状态。让代理为每个请求打开一个新连接是完全可以接受的(也是建议的)。指定close作为这些头的值会提醒web服务器，您的代理打算在第一次请求/响应交换后关闭连接。除了Host头，你的代理应该忽略上面描述的请求头的值;相反，您的代理应该始终发送此文档指定的头。为方便起见，在proxy.c中将描述的User-Agent头的值作为字符串常量提供给您。

最后，如果浏览器作为HTTP请求的一部分发送任何额外的请求头，您的代理应该不加修改地转发它们。

4.3端口号

这个实验有两类重要的端口号:HTTP请求端口和您的代理侦听端口。

HTTP请求端口是HTTP请求的URL中的一个可选字段。也就是说，URL的形式可能是http://www.cmu.edu:8080/hub/index.html，在这种情况下，您的代理应该通过端口8080连接到主机www.cmu.edu，而不是默认的HTTP端口80。无论URL中是否包含端口号，您的代理都必须正常工作。

侦听端口是您的代理应该侦听传入连接的端口。您的代理应该接受指定代理侦听端口号的命令行参数。例如，使用以下命令，您的代理应该侦听端口12345上的连接:

linux> ./proxy 12345

每次运行代理测试时，都必须提供端口号。您可以选择任何非特权端口(大于1024小于65,536)，只要它没有被其他进程使用