**HTTP协议分析与编程**

姓名：于朝阳 学号：1410636 专业：计算机科学与技术

1. **实验要求**

分析HTTP协议，编写一个简单的Web服务器。

1）利用Wireshark捕获数据包，分析一个浏览器与Web服务器的交互过程。

2）编写一个简单的Web服务器。a）实现的服务器应能与标准的浏览器（如IE浏览器）进行简单的交互；b）记录浏览器与服务的交互过程；c）利用HTML语音编写网页（可以不包括图像、声音等多媒体文件），浏览器可通过编写的Web服务器正常访问该网页；d）支持多用户并发访问；e）可以采用C/C++或Java语音进行实现。

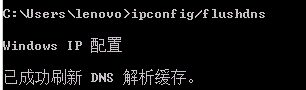
3）选做：扩展编写的简单Web服务器，使浏览器能够浏览Web上存储的图像。

4）提交报告，对以上内容进行总结和分析（包括程序的结构和说明）。

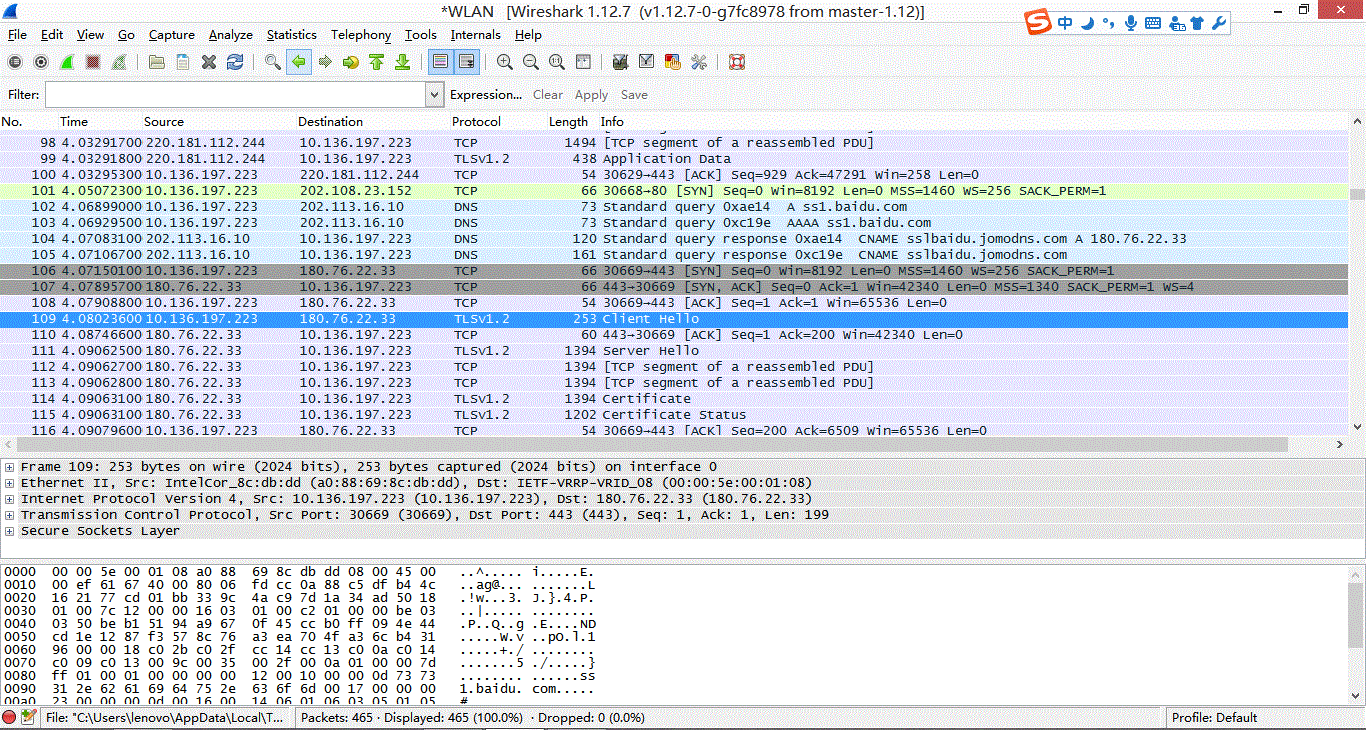
1. **分析数据包**

用Wireshark 捕获数据包

1. 运行wireshark，选择网卡。
2. 将浏览器删除所有的Cookie、临时文件、表单数据以及浏览记录等等，刷新本地DNS。关闭可能的请求网络文件。



1. 开始抓包在浏览器访问www.baidu.com。



1. 分析抓包过程：

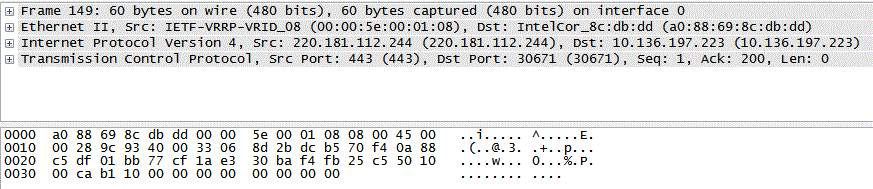
任意选取一个TCP包，可以看到第一行为Frame表示选中的包（一帧），扩展后悔得到更详细的信息，在后面会有具体分析。

选取的包的结构，是按照TCP/IP体系结构的四层模型由下到上来表示的。

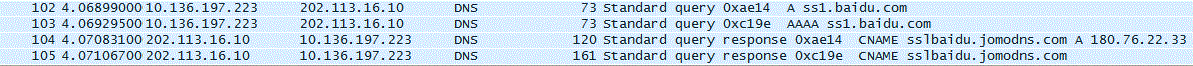
首先是主机-网络层其中包括以太网、令牌环网、无线局域网等，即图中显示的Ethernet 层之后是互联层，即图中的Internet Protocol层

接着是传输层，即图中的TCP层

图中未显示应用层数据，因为这个包是一个TCP层类型包，应用层的数据被封装在里面。



（1）DNS包对www.baidu.com进行翻译，将它翻译为域名所对应的的IP。可以看到本地主机的IP 为202.113.16.10，网址被起名为A，由这个地址像DNS服务器发送请求返回了www.baidu.com的地址180.76.22.33



（2）接下来由TCP采用三次握手的方式进行连接：

第一次握手：建立连接时，客户端发送SYN包(Seq=0)到服务器，并进入SYN\_SEND状态，等待服务器确认。

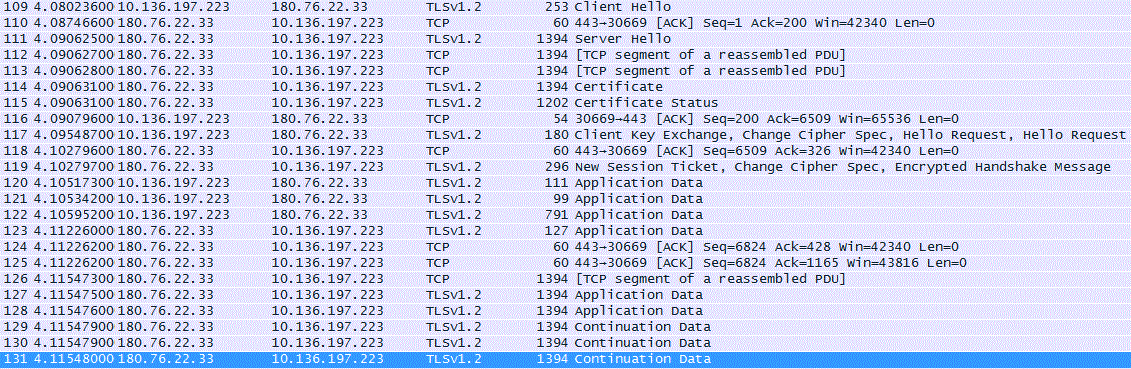
第二次握手：服务器一开始一直监听处在LISTEN状态。服务器收到客户端发来的SYN包，确认客户的SYN发送ACK(Ack=1)，同时也向客户端发送一个SYN包(由图可知Seq=0)，即SYN+ACK包，此时服务器进入SYN\_RECV状态。

第三次握手：客户端收到服务器的SYN＋ACK包，向服务器发送确认包ACK(由图Seq=1)，此包发送完毕，客户端和服务器进入ESTABLISHED状态，完成三次握手。

握手

1. 进行数据交换：

以下截取部分客户端与服务器之间的交换数据的数据包



可以看到图中出现了[TCP segment of a reassembled PDU]：

主机响应客户端的一个请求的时候有可能会回应一个很大的数据，当这些数据超过了TCP的网络传输数据最大值（MSS）的时候，主机发送多个数据包，这里是对于TCP层的分段并不是分片。wireshark里用“Fragmented IP protocol”来标识分片，是IP 层的。这些数据的seq与ACK号是一样的。

（4）关闭TCP链接通过四次挥手进行；

第一次挥手：客户端从ESTABLISHED状态开始关闭应用程序进程，向服务器发送FIN（Seq=1165），关闭与服务器之间的数据传送，然后客户端变为FIN\_WAIT\_1状态。  
 第二次挥手： 服务器从ESTABLISHED状态开始，收到FIN并向客户端发送一个ACK(Ack=1166)，同时向自己对应的进程发送一个文件结束符EOF，状态变为CLOSE\_WAIT。这时，客户端接到ACK后不进行发送，状态变为FIN\_WAIT\_2；  
 第三次挥手： 服务器关闭应用程序进程并关闭与客户端之间的连接，向客户端发送FIN(Seq=16602)，状态变为LAST\_ACK；  
 第四次挥手： 客户端接到FIN后状态变为TIME\_WAIT，并向服务器发出这个FIN的ACK(Ack=16603)确认。之后，客户端等待30秒状态变为CLOSED，服务器接受到这个ACK后不发送进入CLOSED。

综合看来：四次挥手的过程分别是客户端到服务器（FIN）、服务器到客户端（ACK）、服务器到客户端（FIN）、客户端到服务器（ACK）.

捕获到了三个数据包，其中，第二个由服务器到客户端的FIN 和ACK被连在一起发送出来，截图如下：（由于网页中存在其他数据，并且在关闭过程中存在等待时延，所以捕获的到数据包前面的frame号并不是依次排列的)

挥手1

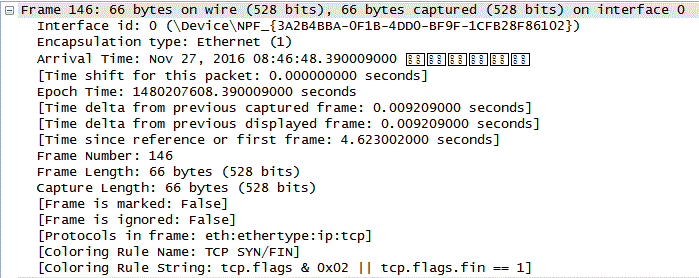
挥手3

挥手2

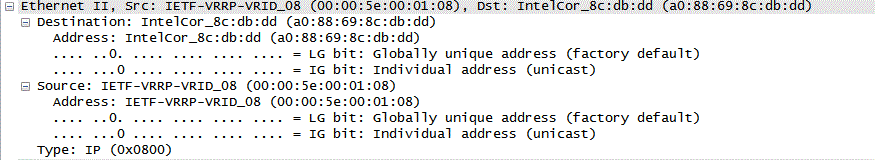
1. 解析数据包：

可以对任意一个包进行一下解析;比如选择一个TCP数据包：

Frame表示一帧，从图中可以看出这里面包含了包的到达时间，抓取间隔时间，包的标号，包的大小，是否标记、忽略等信息。

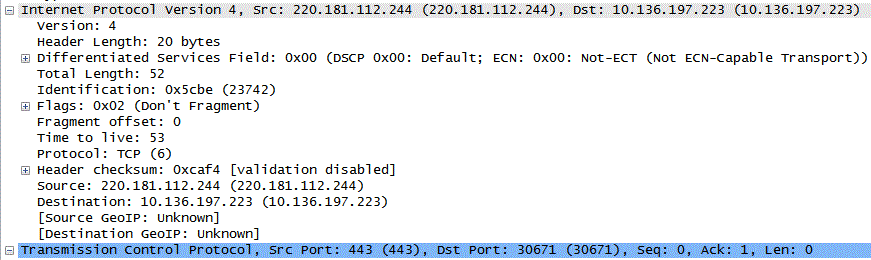


Ethernet层里面包含了目的MAC和源MAC。

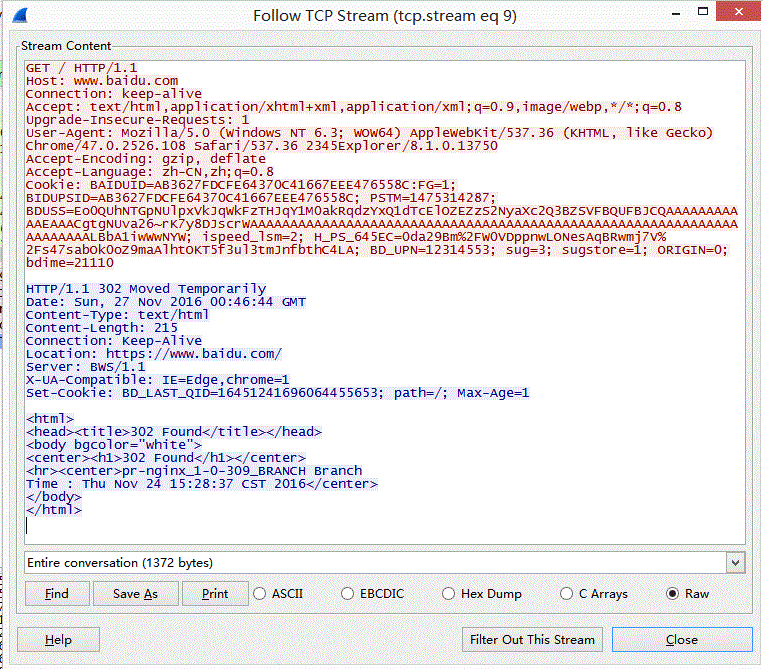


IP层里面包含的信息有版本、源和目的IP、报头长度、采用的运输层协议等信息。

TCP层里边指明了源端口号和目的端口号



还可以看一个HTTP包具体解析：可以读出浏HTTP协议版本、服务器、浏览器版本、时间、请求的内容等信息。



**第一部分是HTTP请求报文：**

GET HTTP/1.1是请求行，GET是方法、HTTP/1.1是版本号

下面的是首部行：

Host:www.baidu.com定义了目标所在的主机即百度服务器

Connection:keep-alive表示发送完请求对象之后并不关闭连接

User-agent:Mozilla/5.0显示了浏览器版本Netscape浏览器

Acccept:text/html,application之后跟着要接受的文件类型

Accept-language:zh-CN,zh;表示想得到对象的语言版本即中文

Cookie:cookie内容

**第二部分是HTTP响应报文**

HTTP/1.1 302 Moved Temporarily 是初始状态行

HTTP/1.1指明协议版本 302表示重定向，被暂时性转移

下面的是首部行：

Date:sun 27 Nov 2016 00:46:44 GMT指明服务器产生并发送该响应报文的日期和时间，并不是对象创建或者最后修改的时间，而是服务器从他的文件系统中检索到该对象、插入到响应报文并发送该响应报文的时间。

Connect-Type:text/html指明了实体主体中的对象是text/html格式的文本。

Connect-length:215指明被发送对象的字节数。

Location:http://www.baidu.com/

Server:指明该报文由一个Apache Web服务器产生的

X-UA-ompatible:体现了IE文档关于兼容格式的定义

Set-Cookie:设置cookie内容

**第三部分就是html文档的内容**

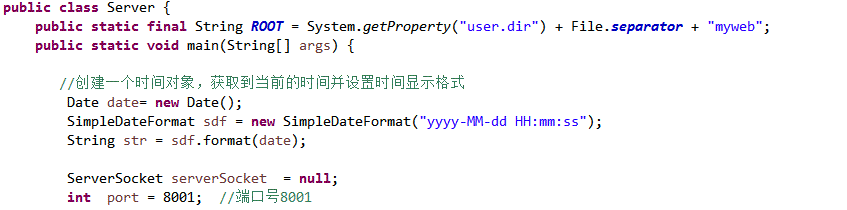
其中设置了网页头部302Found,背景颜色白色、时间等内容。

1. **Web服务器程序结构与说明**

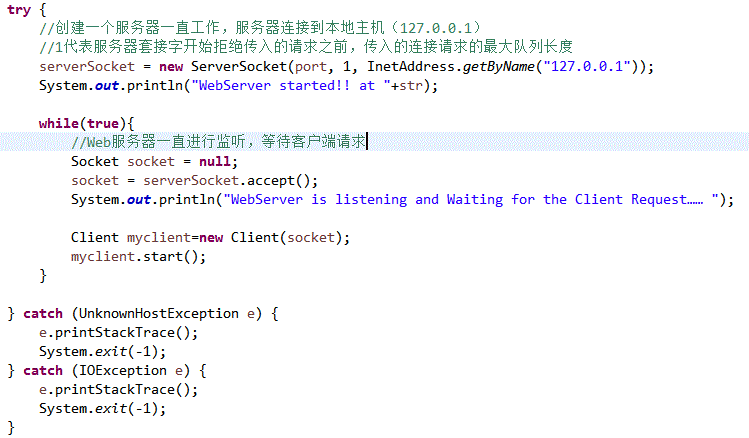
本程序用JAVA语言编写

建立四个class：Server\Request\Response\Client

1. **Server:**首先定义一个ROOT文件夹，在当前工作目录中myweb文件夹下存放静态资源，用来相应客户端的请求。创建ServerSocket 对象，定义端口号为8001。



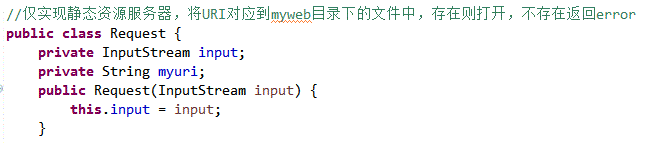
将服务器的IP定义为本机，根据之间定义的端口号和本机IP创建服务器，并让服务器一直一直监听，等待客户端的请求。Start客户端的线程（支持多用户并发访问）。



**Request:**

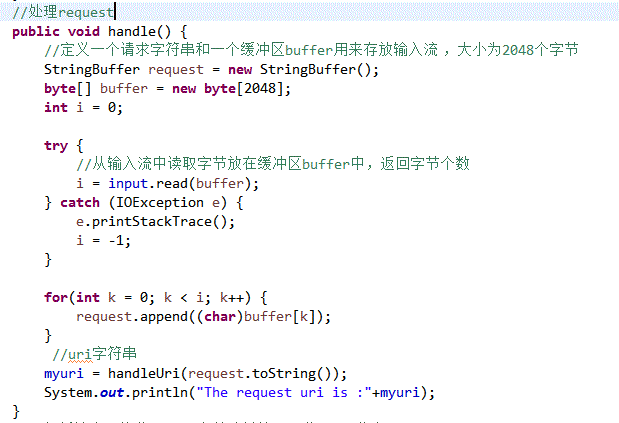
主要思想：解析出URI对应到相应的myweb目录下的文件中。

Request构造函数：



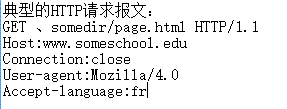
处理Request:

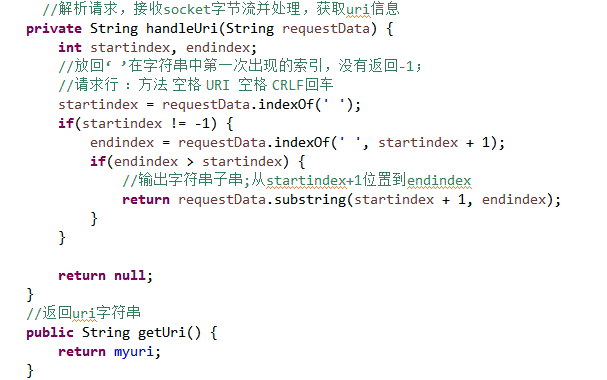
handle函数主要是讲请求从输入流放入缓冲区中，将其转化成字符串以便解析。



handleUri用来将从输入流字符串中得到uri信息：

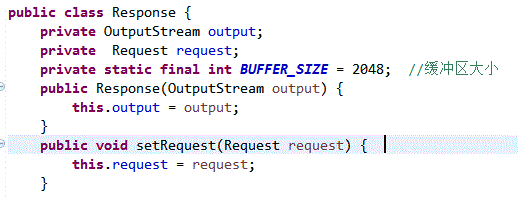
典型的HTTP请求报文格式：

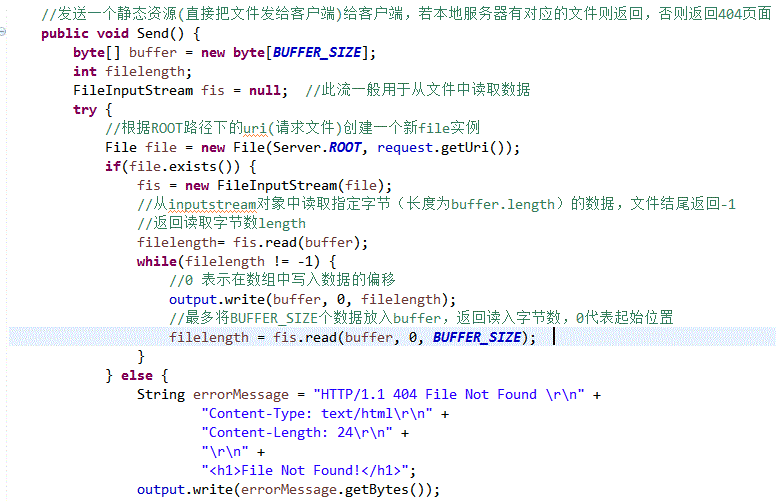


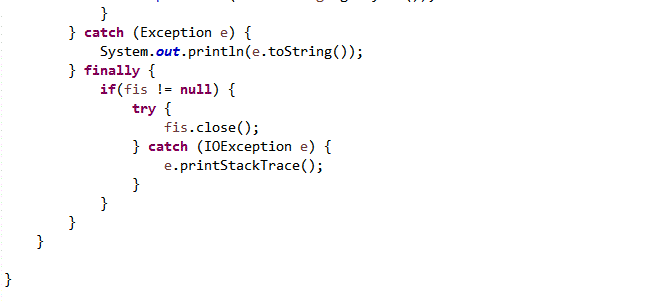


**Response:**

服务器根据Request进行处理查询本地服务器是否有相应的文件，如果没有返回404 Not Found 页面，如果有的话利用FileInputStream从文件中读取数据放入缓冲区，并将响应返回给客户端。

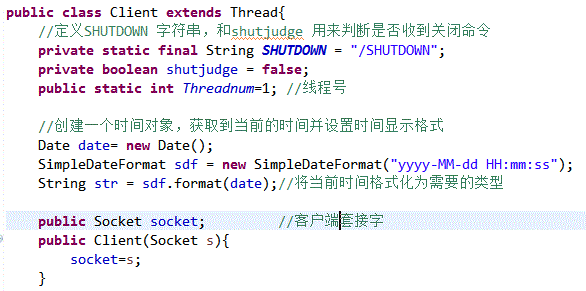


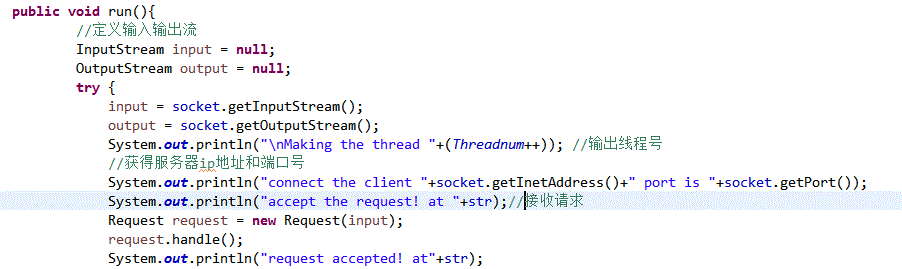


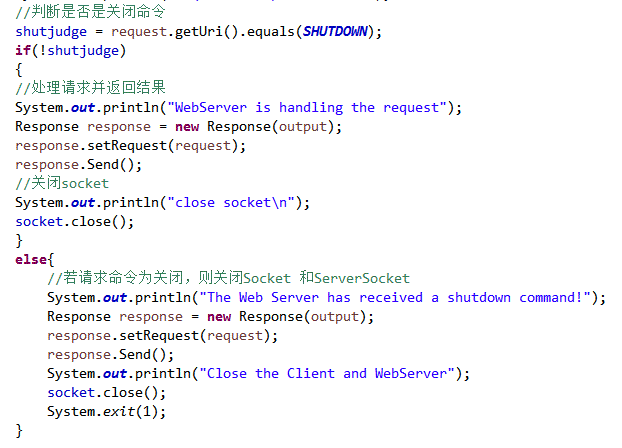


**Client:**

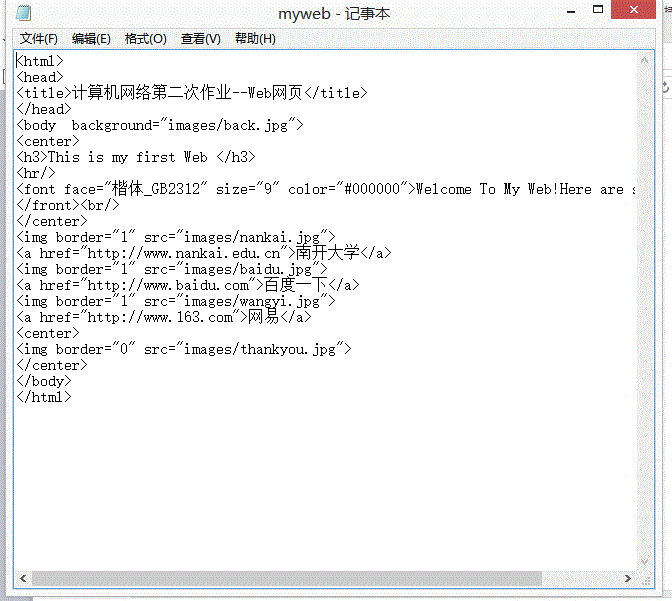
继承Thread类，创建多线程。创建一个Socket套接字，获得每一个线程客户端的IP和端口号，创建输入输出流用来作为Request和Response的参数。运行每一个线程的时候，先用Request类中函数接收到请求进行处理，判断命令是不是shutdown如果是的话要关闭服务器，否则利用Response处理请求，并响应返回给客户端。在每一个线程结束之后都要关闭客户端。







**编写的html文档：**



1. **程序演示结果**

**启动服务器**

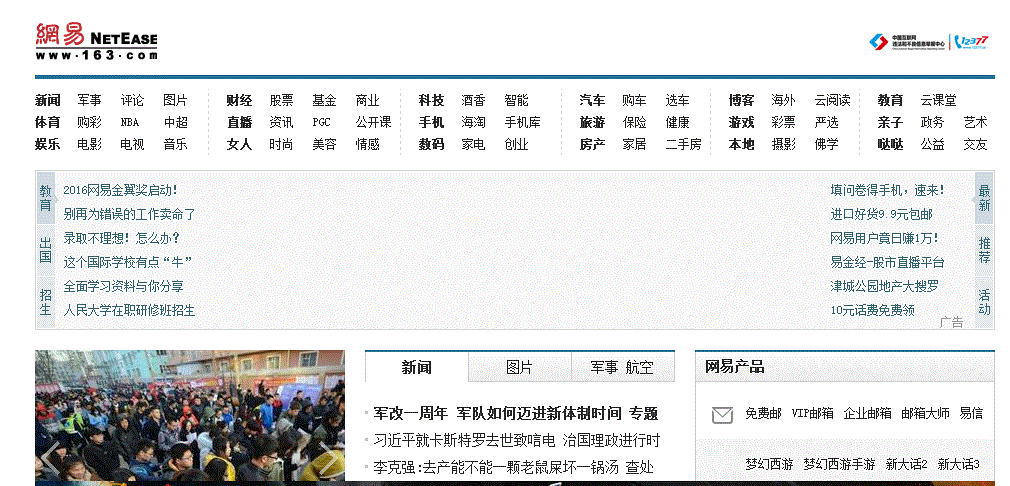
start

**在浏览器中输入网址**

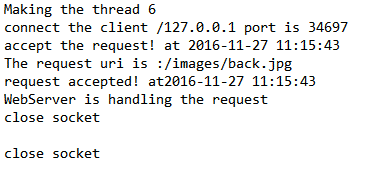
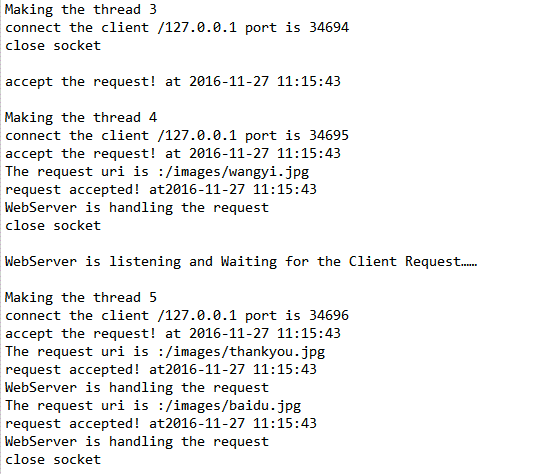
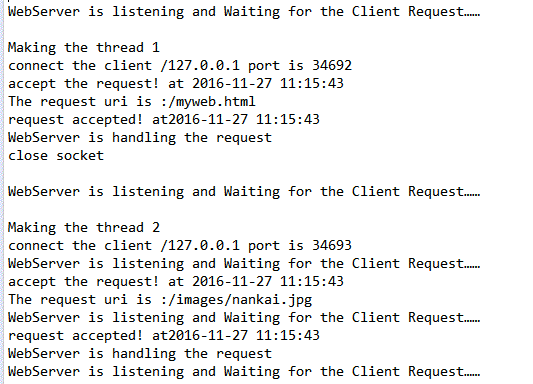
网址

**响应网页**：（包含文字、图片和链接，链接可以分别链接到南开、百度、网易的首页）





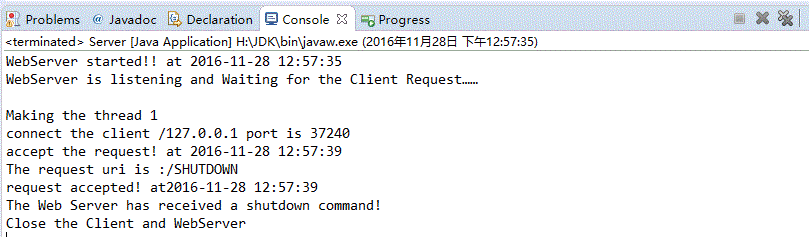
**控制台打印出浏览器与Web服务器的交互过程：**



如果是SHUTDOWN 命令的话：

SHUT

客户端和服务器都被关闭



**同时，其他主机也可以访问本机的服务器，将服务器端口号进行修改：**

本机

运行结果：

