SSN COLLEGE OF ENGINEERING, KALAVAKKAM
(An Autonomous Institution, Affiliated to Anna University, Chennai)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

NETWORKS LAB
EXERCISE 7

Name: Jayannthan P T   *Dept: CSE 'A'*   *Roll No.: 205001049*

HTTP web client program to download the webpage TCP

Aim:

Implement a HTTP web client program to download the webpage using C socket programming.

Algorithm:

1. Read the name of the server as command line argument
(Ex: www.ssn.edu.in/wp-content/uploads/2021/02/NIRF-2021-SSNCE-ENGG.pdf)
2. Get the address of the server using gethostbyname() that returns the pointer to network data structure for given host.
3. Create a TCP socket using socket().
4. Connect to remote server.
5. Send request using a GET /path/filename HTTP/1.1\r\n request using either send() or write().
(GET /wp-content/uploads/2021/02/NIRF-2021-SSNCE-ENGG.pdf HTTP/1.1\r\nHost: www.ssn.edu.in \r\n\r\n)
6. Receive the response using either recv() or read().
7. Parse the response to find out if the request succeeded and what format the file data is being sent as.
8. Receive the file data, if present, using either recv() or read() and write the downloaded page into file under a different name in a local folder.
9. Close the socket and the file.

Code:

```c
#include<stdio.h>
#include<stdlib.h>
#include<string.h>
#include<unistd.h>
#include<fcntl.h>
#include<netdb.h>
#include<sys/types.h>
#include<sys/socket.h>
#include<arpa/inet.h>
#define SERVER_TCP_PORT 80
int main(int argc, char * argv[]) {
    int sockfd, len;
    char hostname[20], path[100], request[120], response[1024], newfile[30];
    struct sockaddr_in serveraddr;
    if (argc == 1)
    {
        printf("Enter URL in command line argument\n");
        return 0;
    }
    int i,j;
    for (i = 0; i < strlen(argv[1]); i++)
    {
        if (argv[1][i] != '/')
        {
            hostname[i] = argv[1][i];
        }
        else
        {
            break;
        }
    }
    for (i = i, j = 0; i < strlen(argv[1]); i++, j++)
    {
        path[j] = argv[1][i];
    }
    path[j] = '\0';
    strcpy(request, "GET ");
    strcat(request, path);
    strcat(request, " HTTP/1.1\r\nHost: ");
    strcat(request, hostname);
    strcat(request, "\r\n\r\n");
    struct hostent * host = gethostbyname(hostname);
    sockfd = socket(AF_INET, SOCK_STREAM, 0);
    serveraddr.sin_family = AF_INET;
    serveraddr.sin_addr = * ((struct in_addr * ) host -> h_addr);
    serveraddr.sin_port = htons(SERVER_TCP_PORT);
    connect(sockfd, (struct sockaddr * ) & serveraddr, sizeof(serveraddr));
    write(sockfd, request, strlen(request));
```

```c
        len = read(sockfd, response, 1024);
        close(sockfd);
        printf("Enter file name to save: ");
        scanf("%s", newfile);
        FILE * fd = fopen(newfile, "w+");
        fwrite(response, 1, len, fd);
        fclose(fd);
        printf("Downloaded file saved under %s\n", newfile);
        return 0;
}
```

Output:

```
root@spl13:~/Desktop/Jayannthan# gcc A7.c
root@spl13:~/Desktop/Jayannthan# ./a.out www.ssn.edu.in/wp-content/uploads/2021/02/NIRF-2021-SSNCE-ENGG.pdf
Enter file name to save: pdfhtml
Downloaded file saved under pdfhtml
root@spl13:~/Desktop/Jayannthan# ls pdfhtml
pdfhtml
root@spl13:~/Desktop/Jayannthan# cat pdfhtml
HTTP/1.1 301 Moved Permanently
Date: Fri, 14 Oct 2022 05:55:06 GMT
Server: Apache
Location: https://www.ssn.edu.in/wp-content/uploads/2021/02/NIRF-2021-SSNCE-ENGG.pdf
Content-Length: 282
Connection: close
Content-Type: text/html; charset=iso-8859-1

<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 2.0//EN">
<html><head>
<title>301 Moved Permanently</title>
</head><body>
<h1>Moved Permanently</h1>
<p>The document has moved <a href="https://www.ssn.edu.in/wp-content/uploads/2021/02/NIRF-2021-SSNCE-ENGG.pdf">here</a>.</p>
</body></html>
```

Learning outcome:

Learnt to implement a HTTP web client program