

```
1: // $RCSfile: httpget.c,v $$Revision: 1.1 $
2:
3: //
4: // httpget - get a web page from some host
5: //
6: // usage:
7: //      httpget [-port] http_host pathname
8: //
9: // Default port is 80.  Gets the given page from the given host.
10: //
11:
12: #define DEFAULT_PORT 80
13:
14: #include <errno.h>
15: #include <libgen.h>
16: #include <netdb.h>
17: #include <netinet/in.h>
18: #include <stdio.h>
19: #include <stdlib.h>
20: #include <string.h>
21: #include <sys/socket.h>
22: #include <sys/types.h>
23: #include <unistd.h>
24:
25: char *exec_name = NULL;
26:
27: void usage() {
28:     fprintf (stderr, "Usage: %s [-port] [host [path]]\n", exec_name);
29:     exit (EXIT_FAILURE);
30: }
31:
32: void print_command (char **argv) {
33:     printf ("\nCommand:");
34:     while (*argv) printf (" %s", *argv++);
35:     printf ("\n");
36: }
37:
38: void crash (const char *object, const char *message) {
39:     fflush (stdout);
40:     fprintf (stderr, "%s: %s: %s\n", exec_name, object, message);
41:     fflush (stderr);
42:     exit (EXIT_FAILURE);
43: }
44:
45: void print_host_port (char *http_host, struct hostent *hostp,
46:                      int http_port) {
47:     printf ("host = %s = %s =", http_host, hostp->h_name);
48:     for (int i = 0; i < 4; i++) {
49:         printf ("%c%d", i == 0 ? ' ' : '.',
50:                 (unsigned char) hostp->h_addr[i]);
51:     }
52:     printf ("\nport = %d\n", http_port);
53: }
54:
```

```
55:
56: int main (int argc, char *argv[]) {
57:     print_command (argv);
58:     exec_name = basename ((argc--, *argv++));
59:     int http_port = DEFAULT_PORT;
60:     if (argc > 0 && **argv == '-') http_port = atoi ((argc--, *argv++));
61:     char *http_host = argc > 0 ? (argc--, *argv++) : "localhost";
62:     char *http_path = argc > 0 ? (argc--, *argv++) : "/";
63:     if (argc != 0) usage();
64:
65:     // get inet address of host and print it
66:     struct hostent *hostp = gethostbyname (http_host);
67:     if (hostp == 0) crash (http_host, hstrerror (h_errno));
68:     print_host_port (http_host, hostp, http_port);
69:
70:     // create a socket to connect to other machine
71:     int sockfd = socket (AF_INET, SOCK_STREAM, 0);
72:     if (sockfd < 0) crash ("socket()", strerror (errno));
73:
74:     // create a struct sockaddr and connect it to the socket
75:     struct sockaddr_in sock;
76:     sock.sin_family = AF_INET;
77:     sock.sin_port = htons (http_port);
78:     sock.sin_addr = *(struct in_addr *) hostp->h_addr;
79:     int rc = connect (sockfd, (struct sockaddr *) &sock, sizeof sock);
80:     if (rc < 0) crash ("connect()", strerror (errno));
81:
82:     // read info from the remote and convert to host format
83:     char buffer[4096];
84:     int nbytes = sprintf (buffer, "GET %s\n", http_path);
85:     printf ("request: %s", buffer);
86:     write (sockfd, buffer, nbytes);
87:     for(;;){
88:         int nbytes = read (sockfd, &buffer, sizeof buffer - 1);
89:         if (nbytes == 0) break;
90:         buffer[nbytes] = '\0';
91:         printf ("%s", buffer);
92:     }
93:
94:     // all done
95:     close (sockfd);
96:     return EXIT_SUCCESS;
97: }
98:
99: //TEST// echo $0: Running httpget tests: >httpget.out
100: //TEST// ./httpget >>httpget.out 2>&1
101: //TEST// ./httpget Unknown.Host >>httpget.out 2>&1
102: //TEST// ./httpget soe.ucsc.edu >>httpget.out 2>&1
103: //TEST// ./httpget www2.ucsc.edu /courses/cmpls109-wm/ >>httpget.out 2>&1
104: //TEST// mkpspdf httpget.ps httpget.c* httpget.out
105:
```

[illegible]

```
1: bash: Running httpget tests:
2:
3: Command: ./httpget
4: host = localhost = localhost = 127.0.0.1
5: port = 80
6: httpget: connect(): Connection refused
7:
8: Command: ./httpget Unknown.Host
9: httpget: Unknown.Host: Unknown host
10:
11: Command: ./httpget soe.ucsc.edu
12: host = soe.ucsc.edu = soe.ucsc.edu = 128.114.48.62
13: port = 80
14: request: GET /
15: <!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 2.0//EN">
16: <HTML><HEAD>
17: <TITLE>302 Found</TITLE>
18: </HEAD><BODY>
19: <H1>Found</H1>
20: The document has moved <A HREF="http://www.soe.ucsc.edu/">here</A>.<P>
21: </BODY></HTML>
22:
23: Command: ./httpget www2.ucsc.edu /courses/cmps109-wm/
24: host = www2.ucsc.edu = www2-vip72.ucsc.edu = 128.114.119.72
25: port = 80
26: request: GET /courses/cmps109-wm/
27: <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 3.2 Final//EN">
28: <html>
29: <head>
30: <title>Index of /courses/cmps109-wm</title>
31: </head>
32: <body>
33: <h1>Index of /courses/cmps109-wm</h1>
34: <table><tr><th></th><th><a href="?C=N;O=D">Name</a></th><th><a href="?C=M;O=A">Last modified</a></th><th><a href="?C=S;O=A">Size</a></th><th><a href="?C=D;O=A">Description</a></th></tr><tr><th colspan="5"><hr></th></tr>
35: <tr><td valign="top"></td><td><a href="/courses/">Parent Directory</a></td><td align="right"> - </td></tr>
36: <tr><td valign="top"></td><td><a href=".:/">.:</a></td><td align="right">
12-Mar-2013 19:27 </td><td align="right"> - </td></tr>
37: <tr><th colspan="5"><hr></th></tr>
38: </table>
39: <PRE><A HREF=:>
40:
41: There is no web page for this course, but instead, a Unix
42: directory structure. The proper way to access it is by using
43: ssh to log into a IC Solaris host (unix.ic.ucsc.edu). However,
44: if you actually do like that silly pointy clicky method of
45: accessing information, the local web server will format the Unix
46: directories as web pages for you.
47:
48: If you are looking in this directory properly, i.e., via Unix
49: commands, there is nothing of interest here. Just ``cd ..``.
50: If you are a pointy clicky type, click on the colon above or
51: right here on this README to get to the course directory. And
52: don't call it a folder. The web link labelled Parent Directory
53: above is not actually the parent directory but a link to a
54: course web directory.
55:
56: If you are using a web client on a ucsc.edu host, you should see
57: no password request. From off campus, you need to enter a
```

```
58: username, but may leave the password field blank. To discover
59: the username, log into unix.ic and look at the hidden file
60: ``.htpasswd`` in the parent of this directory.
61:
62: $Id: README.html,v 1.1 2008-11-20 15:40:02-08 - - $
63:
```