

The purpose of this program is to replicate wget's functionality by using write() and read() in C.

Using the website example.com as an example, wget can be replicated, after running the **make** command, by entering the command **bin/myweb www.example.com 93.184.216.34:80/index.html**. The port is optional.

The program's body will be placed in the output.dat file. Appending **-h** to the end of the command will instead post the webpage's header to stdout/the terminal.

Shortcomings:

1. Sometimes the program does not always return the exact same header as wget. (Ignoring time differences.) See tests 1 and 2.
2. softwareqatest.com does not work, it hangs forever on read.
3. When an incorrect port is entered, the program hangs for 2 minutes before terminating gracefully.

Tests (Due to me copying and pasting directly from VS Code, all wget header responses are ahead by one tab):

Test 1 - example.com, basic test

wget 93.184.216.34:80/index.html --header="Host: www.example.com" (left) vs **bin/myweb www.example.com 93.184.216.34:80/index.html** (right)

The two texts are identical!

Edit texts ...

Switch texts

Compare!

Clear all

```
<!doctype html>
<html>
<head>
  <title>Example Domain</title>

  <meta charset="utf-8" />
  <meta http-equiv="Content-type" content="text/html;
charset=utf-8" />
  <meta name="viewport" content="width=device-width, initial-
scale=1" />
  <style type="text/css">
    body {
      background-color: #f0f0f2;
      margin: 0;
      padding: 0;
      font-family: -apple-system, system-ui, BlinkMacSystemFont,
"Segoe UI", "Open Sans", "Helvetica Neue", Helvetica, Arial, sans-
serif;
```

```
<!doctype html>
<html>
<head>
  <title>Example Domain</title>

  <meta charset="utf-8" />
  <meta http-equiv="Content-type" content="text/html;
charset=utf-8" />
  <meta name="viewport" content="width=device-width, initial-
scale=1" />
  <style type="text/css">
    body {
      background-color: #f0f0f2;
      margin: 0;
      padding: 0;
      font-family: -apple-system, system-ui, BlinkMacSystemFont,
"Segoe UI", "Open Sans", "Helvetica Neue", Helvetica, Arial, sans-
serif;
```

wget 93.184.216.34:80/index.html --header="Host: www.example.com" --server-response
--spider (left) vs **bin/myweb www.example.com 93.184.216.34:80/index.html -h** (right)

1 HTTP/1.1 200 OK
2 Accept-Ranges: bytes
3 Age: 91196
4 Cache-Control: max-age=604800
5 Content-Type: text/html; charset=UTF-8
6 Date: Fri, 14 Jan 2022 23:56:05 GMT
7 Etag: "3147526947"
8 Expires: Fri, 21 Jan 2022 23:56:05 GMT
9 Last-Modified: Thu, 17 Oct 2019 07:18:26 GMT
10 Server: ECS (oxr/8321)
11 X-Cache: HIT
12 Content-Length: 1256

1 HTTP/1.1 200 OK
2 Accept-Ranges: bytes
3 Age: 91220
4 Cache-Control: max-age=604800
5 Content-Type: text/html; charset=UTF-8
6 Date: Fri, 14 Jan 2022 23:56:29 GMT
7 Etag: "3147526947"
8 Expires: Fri, 21 Jan 2022 23:56:29 GMT
9 Last-Modified: Thu, 17 Oct 2019 07:18:26 GMT
10 Server: ECS (oxr/8321)
11 Vary: Accept-Encoding
12 X-Cache: HIT
13 Content-Length: 1256

Test 2 - pudim.com.br, basic test

wget 54.207.20.104:80/index.html --header="Host: www.pudim.com.br" (left) vs
bin/myweb www.pudim.com.br 54.207.20.104:80/index.html (right)

The two texts are identical!

Edit texts ...

Switch texts

Compare!

Clear all

```
<html>
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
  <title>Pudim</title>
  <link rel="stylesheet" href="estilo.css">
</head>
<body>
<div>
  <div class="container">
    <div class="image">
      
    </div>
    <div class="email">
      <a
href="mailto:pudim@pudim.com.br">pudim@pudim.com.br</a>
    </div>
  </div>
</div>
</div>
```

```
<html>
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
  <title>Pudim</title>
  <link rel="stylesheet" href="estilo.css">
</head>
<body>
<div>
  <div class="container">
    <div class="image">
      
    </div>
    <div class="email">
      <a
href="mailto:pudim@pudim.com.br">pudim@pudim.com.br</a>
    </div>
  </div>
</div>
```

```
wget 54.207.20.104:80/index.html --header="Host: www.pudim.com.br" --server-response
```

--spider (left) vs bin/myweb www.pudim.com.br 54.207.20.104:80/index.html -h (right)

```
1 HTTP/1.1 200 OK
2 Date: Sat, 15 Jan 2022 00:03:45 GMT
3 Server: Apache/2.4.34 (Amazon) OpenSSL/1.0.2k-fips
  PHP/5.5.38
4 Last-Modified: Wed, 23 Dec 2015 01:18:20 GMT
5 ETag: "353-527867f65e8ad"
6 Accept-Ranges: bytes
7 Content-Length: 851
8 Keep-Alive: timeout=5, max=100
9 Connection: Keep-Alive
10 Content-Type: text/html; charset=UTF-8
```

Test 3 - mcafeesites.com, basic test

wget 161.69.49.22:80/index.html --header="Host: www.testingmcafeesites.com" (left) vs
bin/myweb www.testingmcafeesites.com 161.69.49.22:80/index.html (right)

Email this comparison

Clear all

<div>1</div> <div>2 <html></div> <div>3 <head></div> <div>4 <title>URL for testing.</title></div> <div>5 </head></div> <div>6 <body></div> <div>7 <code></div> <div>8 <A</div> <div>9 HREF="http://www.testingmcafeesites.com/index.html">http://</div> <div>10 /www.testingmcafeesites.com/index.html
</div> <div>11
</div> <div>12 This is an index url which gives</div> <div>13 an overview of the different test urls available.
</div> <div>14
</div> <div>15 <A</div> <div>16 HREF="http://www.testingmcafeesites.com/testcat_ac.html">h</div> <div>17 ttp://www.testingmcafeesites.com/testcat_ac.html
</div> <div>18 -This is an example URL which</div> <div>19 should be categorized as an art/culture website with a</div>	<div>1</div> <div>2 <html></div> <div>3 <head></div> <div>4 <title>URL for testing.</title></div> <div>5 </head></div> <div>6 <body></div> <div>7 <code></div> <div>8 <A</div> <div>9 HREF="http://www.testingmcafeesites.com/index.html">http://</div> <div>10 /www.testingmcafeesites.com/index.html
</div> <div>11
</div> <div>12 This is an index url which gives</div> <div>13 an overview of the different test urls available.
</div> <div>14
</div> <div>15 <A</div> <div>16 HREF="http://www.testingmcafeesites.com/testcat_ac.html">h</div> <div>17 ttp://www.testingmcafeesites.com/testcat_ac.html
</div> <div>18 -This is an example URL which</div> <div>19 should be categorized as an art/culture website with a</div>
---	---

wget 161.69.49.22:80/index.html --header="Host: www.testingmcafeesites.com"
--server-response --spider (left) vs **bin/myweb www.testingmcafeesites.com**
161.69.49.22:80/index.html -h (right)

<div>1 HTTP/1.1 200 OK</div> <div>2 Content-Type: text/html</div> <div>3 Content-Length: 28634</div> <div>4 Connection: keep-alive</div> <div>5 Last-Modified: Mon, 22 Jun 2020 20:38:18 GMT</div> <div>6 Accept-Ranges: bytes</div> <div>7 ETag: "c7371810d548d61:0"</div> <div>8 Server: Microsoft-IIS/8.5</div> <div>9 X-Powered-By: ASP.NET</div> <div>10 SN: LMINIWEB02</div> <div>11 Date: Sat, 15 Jan 2022 00:11:03 GMT</div>	<div>1 HTTP/1.1 200 OK</div> <div>2 Content-Type: text/html</div> <div>3 Content-Length: 28634</div> <div>4 Connection: keep-alive</div> <div>5 Last-Modified: Mon, 22 Jun 2020 20:38:18 GMT</div> <div>6 Accept-Ranges: bytes</div> <div>7 ETag: "c7371810d548d61:0"</div> <div>8 Server: Microsoft-IIS/8.5</div> <div>9 X-Powered-By: ASP.NET</div> <div>10 SN: LV2MINIWEB01</div> <div>11 Date: Sat, 15 Jan 2022 00:10:34 GMT</div>
--	--

Test 4 - neverssl.com, basic test

wget 52.222.158.89:80/index.html --header="Host: www.neverssl.com" (left) vs **in/myweb www.neverssl.com 52.222.158.89:80/index.html** (right)

The two texts are identical!

Edit texts ... Switch texts Compare! Clear all

```
<html>
  <head>
    <title>NeverSSL - Connecting ... </title>
    <style>
      body {
        font-family: Montserrat, helvetica, arial,
        sans-serif;
        font-size: 16x;
        color: #444444;
        margin: 0;
      }
      h2 {
        font-weight: 700;
        font-size: 1.6em;
      }
    </style>
  </head>
  <body>
    <h2>
      Connecting to 52.222.158.89:80
    </h2>
  </body>
</html>
```

wget 52.222.158.89:80/index.html --header="Host: www.neverssl.com" --server-response
--spider (left) vs **in/myweb www.neverssl.com 52.222.158.89:80/index.html -h** (right)

Email this comparison

1	HTTP/1.1 200 OK	1	HTTP/1.1 200 OK
2	Content-Type: text/html	2	Content-Type: text/html
3	Content-Length: 3961	3	Content-Length: 3961
4	Connection: keep-alive	4	Connection: keep-alive
5	Date: Fri, 14 Jan 2022 04:07:55 GMT	5	Date: Fri, 14 Jan 2022 04:07:55 GMT
6	Last-Modified: Thu, 04 Nov 2021 18:27:33 GMT	6	Last-Modified: Thu, 04 Nov 2021 18:27:33 GMT
7	ETag: "41f0211ff315bbd7e2b6dc2e98143935"	7	ETag: "41f0211ff315bbd7e2b6dc2e98143935"
8	Accept-Ranges: bytes	8	Accept-Ranges: bytes
9	Server: AmazonS3	9	Server: AmazonS3
10	Vary: Accept-Encoding	10	Vary: Accept-Encoding
11	Cache-Control: public, max-age=86400	11	Cache-Control: public, max-age=86400
12	X-Cache: Hit from cloudfront	12	X-Cache: Hit from cloudfront
13	Via: 1.1 127aaaca740f298a4c887357ec047b4.cloudfront.net (CloudFront)	13	Via: 1.1 16a28c0e67da18fa2960e2e414084d76.cloudfront.net (CloudFront)
14	X-Amz-Cf-Pop: CDG52-P2	14	X-Amz-Cf-Pop: CDG52-P2
15	X-Amz-Cf-Id: 134tQVh8GBH-a9TAcimPsVLN1cJaWrzZ-P47A3zfA_h0bIE0tpSxw==	15	X-Amz-Cf-Id: crZCJ8Yvn_0PF3tigYKNeuHGUQGVY3gv6whqJJ_s-TuSJ_DA0kDjw==
16	Age: 72536	16	Age: 72530

Test 5, intentional errors

5.1 Intentional Path Error: **bin/myweb www.neverssl.com 52.222.158.89:80/thishtmlisnotreal**

```
bash-4.2$ bin/myweb www.neverssl.com 52.222.158.89:80/thishtmlisnotreal  
HTTP/1.1 403 Forbidden
```

5.2 Chunked Encoding: **bin/myweb www.google.com 216.58.213.68:80/index.html**

```
bash-4.2$ bin/myweb www.google.com 216.58.213.68:80/index.html  
Chunked encoding is not supported!
```

5.3 Invalid Port: **bin/myweb www.testingmcafeesites.com 161.69.49.22:90/index.html -h**

```
bash-4.2$ bin/myweb www.testingmcafeesites.com 161.69.49.22:90/index.html -h  
Connect error. Check IP and Ports!
```