

情報科学プロジェクト実験レポート課題

S142063 佐藤涼亮

平成 28 年 11 月 13 日

Web クライアント

1 課題の内容

Web ページの取得

1.1 要点

- Web ページを取得するプログラムの作成。(html の内容)
- 配布されたプログラム (cil.cpp) をもとに作成。
- 対象となる Web ページは、その URL をコマンド引数として指定する。
- 取得する Web ページのサイズは予めわからないため、サーバが接続を切るまで繰り返し受信を行う。

2 プログラムの説明

実行時、コマンドライン引数に取得する Web ページの URL(例: `http://www.cc.seikei.ac.jp/index.html`) を指定する。プログラム上で URL を受け取り、ホスト部 (例: `www.cc.seikei.ac.jp`) とパス (例: `/index.html`) に分解し入手する。入手したホスト部の情報をもとに接続先情報の準備を行う。(ポート番号は 80 で固定) その情報をもとに TCP ソケットの作成を行う。作成したソケットを用いて準備したホストへの接続を試みる。接続が成功すれば、HTTP リクエストの GET の要求を入手したパスで指定して HTTP/1.0 でサーバに送信する。メッセージバッファサイズは 128 で固定し、たりない分は繰り返し終わるまで受信を繰り返す。表示する部分は、HTTP リクエストの部分を省いた、空行以降 (html の内容) を表示する。最後まで表示したら、サーバとの接続を遮断し終了する。

2.1 目的

TCP/IP を利用した通信プログラムの作成

2.2 方法

HTTP による Web サーバへの要求 (HTTP リクエスト)

2.3 結果

`http://islay.ci.seikei.ac.jp/` と `http://islay.ci.seikei.ac.jp/~okam/index.html` を引数にした時の実行結果を以下に記す

ソースコード 1: islay.ci.seikei.ac.jp

```
1 <!DOCTYPE html>
2 <html>
3 <head>
4 <meta charset="UTF-8">
5 <title>Welcome to Computer System Labo. </title>
6 </head>
7 <frameset cols="10%,3%,87%" frameborder="0">
8 <frame src="/outline.html" style="background-color:#339933">
9 <frame src="/none.html" scrolling="no" style="background-color:#CC9933">
10 <frame src="/title.html" scrolling="no" style="background-color:#CCCC33">
11 </frameset>
12 <noframes>
13 <body text=#000000 bgcolor=#FFFF00 link=#0000CC vlink=#000080 alink=#0000CC>
14 <src="/outline.html" name="noframe">
15 </noframes>
16 </body>
17 </html>
```

ソースコード 2: islay.ci.seikei.ac.jp

```
1 <!DOCTYPE html>
2 <html>
3 <head>
4 <meta charset="utf-8">
5 <meta HTTP-EQUIV=REFRESH CONTENT="90">
6 <title>Shusuke's Web Page </title>
7
8 <script>
9 function timePhoto(){
10 var m1 = "<img src=\"img/okamoto2011.jpg\">"
11 var m2 = "<img src=\"img/okam.small.png\">"
12 var day = new Date();
13 var year = day.getFullYear();
14 var month = day.getMonth();
15 var date = day.getDate();
16 var hour = day.getHours();
17 var m = day.getMinutes();
18 var mins = ""+m
19 var mm = ["Jan", "Feb", "Mar", "Apr", "May", "Jun",
20           "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"];
21 document.write(date,"/",mm[month],"/",year,"  
");
22 if (m < 10) { mins = "0"+m; }
23 if (hour < 12) { document.write(hour,":",mins,"<br>", m1); }
24 else
25 if (hour <= 12) { document.write("お昼休み...<br>", m2); }
26 else
27 if (hour < 18) { document.write(hour,":",mins,"<br>", m1); }
28 else
29 if (hour <= 18) { document.write("帰宅したい...<br>", m2); }
30 else
31 { document.write(hour,":",mins,"<br>", m1); }
32 }
33
```

```

34 function setBG(){
35   var idx;
36   var c = [0x228b22, 0x8a2be2, 0xff1493, 0x000080,
37           0xdaa520, 0x8b0000, 0x224422];
38   idx = Math.floor(Math.random()*c.length);
39   document.bgColor = c[idx];
40 }
41
42 function setting() {
43   setBG();
44   timePhoto();
45 }
46 </script>
47
48 </head>
49
50 <body bgcolor="#224422" text="#ffffff" link="#ffff00" vlink="#ffffff" alink="0000ff"
51   >
52   <a href="index_e.html">English</A><P>
53   <script> setting() </script>
54
55
56   <font size="10">岡本秀輔/S. Okamoto</font><br>
57   <hr>
58   <ul>
59     <li>所属
60     <a href="http://www.seikei.ac.jp/">成蹊学園</a>
61     <a href="http://www.seikei.ac.jp/university/index.html">成蹊大学</A>
62     <a href="http://www.seikei.ac.jp/university/rikou/">理工学部</A>
63     <a href="http://www.ci.seikei.ac.jp/">情報科学科</A>
64     <a href="http://www.seikei.ac.jp/university/rikou/graduate/index.html">
65     大学院 理工学研究科</a>
66   </li>
67
68   <li>役職 成蹊大学 教授</li>
69
70   <li>研究活動</li>
71   <dl compact>
72     <dt></dt>
73     <dd>・<a href="./info/gyouseki.html">論文等一覧</a>,
74     <a href="./open/">公開ソフトウェア</a>
75   </dd>
76   <dt></dt><dd>・並列処理プログラミング環境</dd>
77   <dt></dt><dd>・<a href="/islay">対話型アニメーションプログラミング</a></dd>
78   <dt></dt><dd>・
79   <a href="http://www.ipsj.or.jp/">情報処理学会</a>,
80   <a href="http://www.ieice.or.jp/">電子情報通信学会</a>(<
81   <a href="http://www.ieice.org/iss/cw/jpn/">
82   サイバーワールド時限研究専門委員会</a>),
83   <a href="http://www.jsst.or.jp/">日本ソフトウェア科学会</a>,
84   </dd>
85   <dt></dt><dd>・
86   <a href="http://computer.org/">IEEE Computer Society</a>,
87   <a href="http://www.acm.org/">ACM</a>,
88   <a href="https://sites.google.com/site/iwssm/">国際ワークショップ WSSM</
89   a>
90   </dd>
91   </dl>
92   <li>担当科目</li>
93   <dl compact>
94     <dt>学 部:</dt>
95     <dd>
96     フレッシュャーズセミナー,
97     <a href="http://islay.ci.seikei.ac.jp/cs/">コンピュータシステム<a>,
98     応用プログラミング,
99     <a href="http://islay.ci.seikei.ac.jp/os/">オペレーティングシステム</a>,
100    <a href="http://islay.ci.seikei.ac.jp/unix-c/">Unix・C プログラミング</a>,

```

```

100     情報科学コース実験,
101     情報科学プロジェクト実験,
102     輪講, 卒業研究</dd>
103     <dt>大学院:</dt>
104     <dd>コンピュータシステム特論, 情報科学特別実験, 他</dd>
105     </dl>
106     <li>資 料 : </li>
107     <dl>
108         <dd><a href="./info/lecture.html">これまでの担当科目</a>,
109         <a href="./open/">過去の講義ノート</a>,
110         <a href="./info/reading.html">輪講のテキスト</a>
111         </dd>
112     </dl>
113
114     <li>教員室 :   1 1 号館 4 階(1413)</li>
115 </ul>
116
117 <center>
118 <table>
119 <tr align=left>
120     <td><a href="http://islay.ci.seikei.ac.jp/">
121         研究室</a></td>
122     <td><a href="./lang.html">
123         並列プログラム</a></td>
124 </tr>
125 </table>
126 </center>
127
128 <!-- href="other.html"その他
129 <td><a href="./event/">
130     イベント</a></td>
131 <dt></dt><dd>・<a href="ILP/"
132     >命令レベル並列プロセッサのための命令セットアーキテクチャ</a></dd>
133 -->
134 <hr>
135 <address>
136     
137 </address>
138
139 </body>
140 </html>

```

2.4 考察

HTTP リクエスト部分の表示を省き、html の内容のみの出力に成功し、期待通りの結果が得られた。

3 感想

今まで、授業でも C++ のプログラムは学んできたが、オフラインで行うことしかなかった。しかし、今回の課題では Web 上のページにアクセスして、Web ページを取得するプログラムを作成した。より高度なプログラムに触れて、いっそうプログラミングへの意識が高まった。このような、プログラムをすることで、以前に他の授業で学んだソケットの話や、クライアントとサーバ間の通信の動きについてより理解を深められた。

「プログラミングの理解を深めるなら C 言語や C++ を勉強すると良い」という話を聞いた事がある。C/C++ の言語は、プログラミング言語の中でも複雑ではあるが、基礎となっていることが多いと言われている。そのため、この言語をより学ぶことでプログラミングの理解を深め、他言語でも活かせるようなプログラマーを目指したい。

4 プログラム

ソースコード 3: report.cpp

```
1 // TCP クライアント
2 #include <iostream>
3 #include <string>
4 #include <unistd.h>
5 #include <sys/types.h>
6 #include <sys/socket.h>
7 #include <netdb.h>
8 #include <netinet/in.h>
9 #include <arpa/inet.h>
10 #include <stdio.h> // for perror()
11 #include <cstring> // for memset()
12 #include <cstdlib> // for exit()
13
14 void msg_proc(int, std::string);
15 void error(const char s[]) { perror(s); exit(1); }
16
17 int main(int argc, char *argv[])
18 {
19     if (argc < 2) {
20         std::cerr << "usage: ./ " << argv[0] << " << URL\n";
21         return 1;
22     }
23
24     // 取得した URL をホストとパスに分解
25     std::string url = argv[1];
26     std::string host = url.substr(url.find_first_not_of("http://"));
27     std::string path = host.substr(host.find_first_of("/"));
28     host = host.substr(0, host.find_first_of("/"));
29
30     // 接続先情報の準備
31     addrinfo hints, *svr;
32     memset(&hints, 0, sizeof(hints));
33     hints.ai_socktype = SOCK_STREAM;
34     hints.ai_family = AF_INET;
35     hints.ai_flags = AI_NUMERICSERV;
36     if (int err = getaddrinfo(host.c_str(), "80", &hints, &svr)) {
37         std::cout << "Error: " << gai_strerror(err) << "\n";
38         return 1;
39     }
40
41     // TCP ソケットの作成
42     int sockfd = socket(svr->ai_family, SOCK_STREAM, 0);
43     if (sockfd < 0) error("socket");
```

```

44
45 // 接続
46 if (connect(sockfd, svr->ai_addr, svr->ai_addrlen) < 0)
47     error("connect");
48 freeaddrinfo(svr); // メモリの開放
49
50 msg_proc(sockfd,path);
51
52 close(sockfd); // サーバ側に切断を伝える
53 return 0;
54 }
55
56 void msg_proc(int sockfd,std::string path)
57 {
58     // 要求の設定
59     std::string s;
60     s = "GET_" + path + "_HTTP/1.0\r\n\r\n";
61
62     if (send(sockfd, s.c_str(), s.size(), 0) < 0)
63         error("send");
64     const int bufsz = 128; // メッセージバッファサイズ
65     char buf[bufsz+1];
66     std::string sbuf;
67     int n=1;
68     bool flag = 0; // 空行の読み込みの判別に使用
69     while(n){
70         n = recv(sockfd, buf, bufsz, 0); // 受信バイト数は最大値のみ指定
71         if (n == -1) error("recv");
72         buf[n] = '\0';
73         sbuf = buf;
74         if(!flag && sbuf.find("\r\n\r\n")!=std::string::npos){ // 空行の検索
75             sbuf=sbuf.substr(sbuf.find("\r\n\r\n")+4); // 存在すれば省き、flagをたてる
76             flag=1;
77         }
78         if(flag) std::cout << sbuf; // 空行以降 (html)の内容の出力
79     }
80     std::cout << std::endl;
81 }

```