

109021115 吳嘉濬 Lab2 readme

1.

Q1: Is it able to be compiled by gcc?

A1: My program is able to be compiled by GCC.

Q2: Does it meet all requirements?

A2: My program meets all requirements.

Q3: How do you handle the response and extract all the hyperlinks?

A3: 利用 while loop 持續讀取資料並存進 receive 直到沒有讀取到新資料(i.e.資料被全部讀取完成)，之後 break 掉 while 迴圈。以下為相關 code segment：

```
while(1){
    char buffer[10240];
    for(int i=0;i<10240;i++)
        buffer[i]='\0';
    valread=read(server_fd,buffer,10240);
    if(valread<=0) break;
    else{
        cnt+=strlen(buffer);
        strncat(receive,buffer,strlen(buffer));
    }
}
```

利用<string.h>函式庫內的 strstr 函式，連續使用兩次 strstr 以篩選 receiver 當中同時出現"<a"和"href="的 string，進而找到所有 hyperlink。

```
char target1[] = "<a";
char *target1_ptr = &target1[0];
char target2[] = "href=";
char *target2_ptr = &target2[0];
```

使用<string.h>函式庫內的 strtok 函式，以一行一個 hyperlink 的形式依序把所有 hyperlink print 出來。

```
//find hyperlinks
int hyper_count = 0;
receive = strstr(receive, target1_ptr);
receive = strstr(receive, target2_ptr);
while (receive != NULL) {
    match_pos[hyper_count++] = receive;
    receive = strstr(receive + strlen(target2), target2_ptr);
}
//print hyperlinks
for (int i = 0; i < hyper_count; i++) {
    hyperlink = strtok(match_pos[i] + strlen(target2), "\\");
    printf("%s\n", hyperlink);
}
```

Q4: What have you learned from this lab?

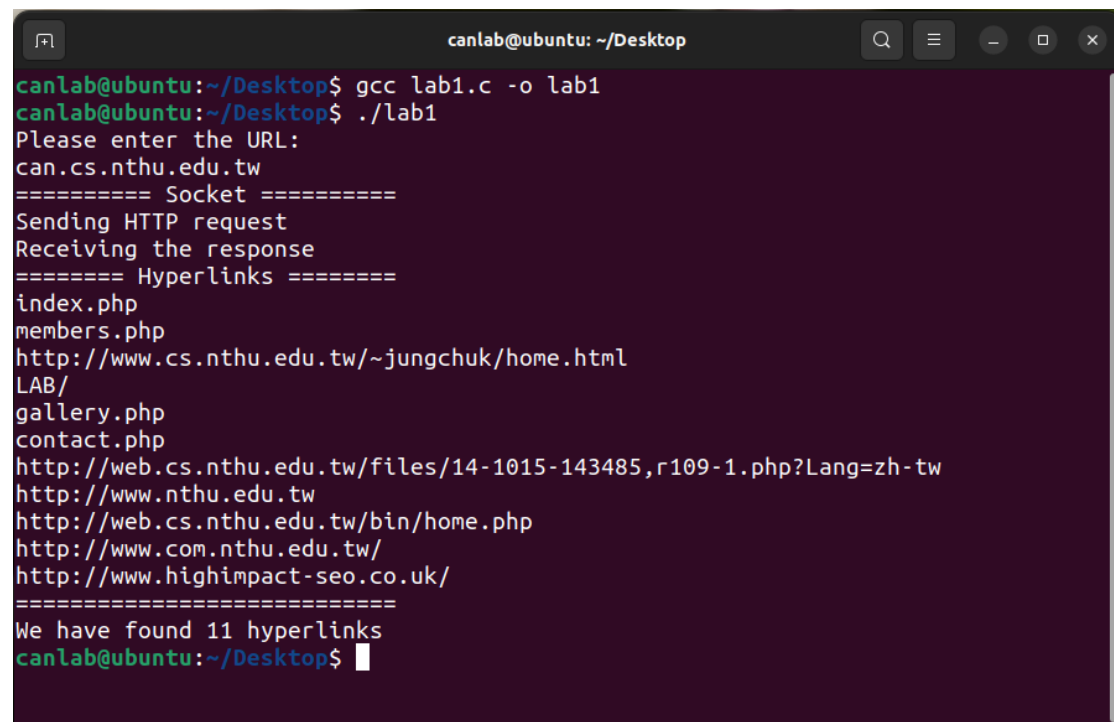
A4: 了解 TCP socket 的運作原理並寫出簡易的 socket programming；學習到如何善用<string.h>函式庫，雖然操作上有些複雜(容易出 bug)，但是卻很實用；熟悉操作基本的 command-line operations。

2.

Screenshot to show how it works.

The following shows 2 screenshots about:

(1) can.cs.nthu.edu.tw

A terminal window titled 'canlab@ubuntu: ~/Desktop' showing the execution of a C program. The user enters the URL 'can.cs.nthu.edu.tw' and the program outputs a list of 11 hyperlinks found on the website. The program's output is as follows:

```
canlab@ubuntu:~/Desktop$ gcc lab1.c -o lab1
canlab@ubuntu:~/Desktop$ ./lab1
Please enter the URL:
can.cs.nthu.edu.tw
===== Socket =====
Sending HTTP request
Receiving the response
===== Hyperlinks =====
index.php
members.php
http://www.cs.nthu.edu.tw/~jungchuk/home.html
LAB/
gallery.php
contact.php
http://web.cs.nthu.edu.tw/files/14-1015-143485,r109-1.php?Lang=zh-tw
http://www.nthu.edu.tw
http://web.cs.nthu.edu.tw/bin/home.php
http://www.com.nthu.edu.tw/
http://www.highimpact-seo.co.uk/
=====
We have found 11 hyperlinks
canlab@ubuntu:~/Desktop$
```

(2) www.cs.nthu.edu.tw/~jungchuk/home.html

```
canlab@ubuntu:~/Desktop$ ./lab1
Please enter the URL:
www.cs.nthu.edu.tw/~jungchuk/home.html
===== Socket =====
Sending HTTP request
Receiving the response
===== Hyperlinks =====
home.html
education.html
publication.html
research.html
project.html
professional.html
teaching.html
lab.html
contact.html
http://www.cs.nthu.edu.tw
http://www.nthu.edu.tw
http://www.com.nthu.edu.tw
http://www.nthu.edu.tw
course/Probability_2021Spring/index.html#adding
course/IntroductionToComputerNetworks_2021Fall/index.html#adding
course/IntroductionToComputerNetworks_2021Fall/index.html#distancelearning
course/BroadbandMobileCommunications_2021Fall/index.html#adding
course/BroadbandMobileCommunications_2021Fall/index.html#distancelearning
research.html
publication.html
research.html
contact.html
research.html
contact.html
research.html
research.html
lab.html
contact.html
http://csim.tca.org.tw/award.aspx
research.html
contact.html
=====
We have found 31 hyperlinks
```