

1. 輸入與輸出

輸入 hostname，也就是一個 http 連線的網址，輸出得到檔案中所有的 hyperlink。

2. 實作方式

- (i). 將 hostname 中分離成 Hostname 和 path
- (ii). 利用 hostname 去尋找得到 ip 位址 (來源見 7. Reference (i).)
- (iii). 建立 TCP socket
- (iv). 使用 socket 和 ip 去和 server 建立 TCP Connection
- (v). send GET http request 給 server
- (vi). 接收 server 的資料到 buffer
- (vii). 因為 buffer 不一定夠大，因此利用演算法將 hyperlink 記錄下來
- (viii). 將所有 hyperlink 印出

3. 將 hyperlink 紀錄之演算法

我們無法事先宣告一個夠大的空間，除非我們和 server 溝通過。此方法為動態接收資料後，存下 url 的同時，記錄一些控制資訊，並設計一個 Linked List 的資料結構，將 url 記錄下來。此章節為 2. 實作方式 (vii). 的詳細介紹。

因為由 hint 知道，url 要能被記錄的前提是

1. 已經出現<a
2. <a 之後還要出現 href=
3. url
4. 以”結束 url
5. 以>結束<a

所以要先找到上述的 1, 2, 3 條件就能記錄下來，然後再根據 4, 5 把紀錄 url 的狀態終止。

Pseudo Code 如下:

```
let c 為接收到的server response之第一個字母;
while c 不是'\0':
    if 已經有 <a 出現了:
        if c是>:
            <a>已經結束了;
        else:
            if 出現了 href=":
                if c是":
                    url的所有字母已經收集完畢，記錄至Linked List中;
                else:
                    把c記錄進一個臨時的buffer中;
```

```

        else:
            一個臨時的buffer會記錄現在是不是href=" ;
            if 發現buffer中的字串加上c時無法成為href="的順序 :
                buffer清空;
    else if < 出現了而已:
        if c 是 a:
            設定<a出現了的控制資訊;
        else:
            設定沒有任何東西出現過的控制資訊;
    else if 沒任何東西出現過，c是< :
        設定 < 出現過的資訊;

    c 往後走一格;

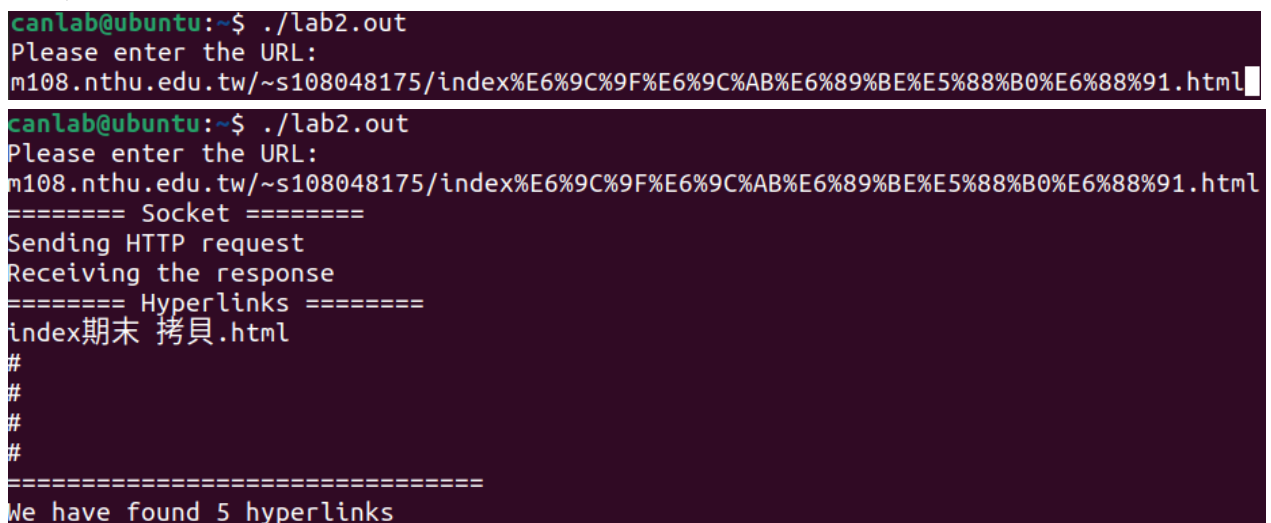
```

4. 使用方式

```
gcc lab.c -o lab2.out
```

```
./lab2.out
```

5. 截圖



```

canlab@ubuntu:~$ ./lab2.out
Please enter the URL:
m108.nthu.edu.tw/~s108048175/index%E6%9C%9F%E6%9C%AB%E6%89%BE%E5%88%B0%E6%88%91.html

canlab@ubuntu:~$ ./lab2.out
Please enter the URL:
m108.nthu.edu.tw/~s108048175/index%E6%9C%9F%E6%9C%AB%E6%89%BE%E5%88%B0%E6%88%91.html
===== Socket =====
Sending HTTP request
Receiving the response
===== Hyperlinks =====
index期末 拷貝.html
#
#
#
#
=====
We have found 5 hyperlinks

```

6. What I Learned

很久沒有寫 c 了，所以花了不少時間處理字串和輸入輸出，如果能重來，應該會想用動態配置記憶體的方式把所有資料存下來，然後一次找出所有的 hyperlink。

雖然寫 lab2 時老師還沒教到 TCP 是 Byte Stream，但覺得寫起來有更了解網路的運作，尤其是可以利用 GET 的方式和別人架設的網站溝通，非常有趣。我有嘗試找了許多 http 網頁，大多是來自系上網站，也可以從中看到很多不同的網站寫法。覺得收穫非常多。

7. Reference

- (i). 將 hostname 轉成 ip: <https://beej-zhtw.netdpi.net/05-system-call-or-bust/5-1-getaddrinfo-start>
- (ii). 傳送 GET request 給 server: <https://stackoverflow.com/questions/30470505/http-request-using-sockets-in-c>