

EECS302002 計算機網路概論

Lab 1

1. Description

Write a program which can show the hyperlinks of a given web page. The main objective is to practice socket programming and try to use HTTP protocol.

2. Requirements

- Input: The users enter the URL of the desired web page without "http://".
e.g. <http://can.cs.nthu.edu.tw/> only need type can.cs.nthu.edu.tw
- Output: Print all hyperlinks and the number of hyperlinks. Note that only `` should be counted while `<link href="xxx">` shouldn't.

3. Example

There are 10 hyperlinks in <http://can.cs.nthu.edu.tw/>.

```
bruce@DESKTOP-DGEBVJ8:/mnt/c/Users/canlab/Desktop$ ./lab1
Enter the hostname: can.cs.nthu.edu.tw
socket: Start send HTTP request
socket: Start read the response
socket: Finish read to buffer
===== Hyperlinks =====
index.php
members.php
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/
```

4. Hint

- Refer to the “[Linux socket tutorial](#)” document for information about linux and socket programming.
- Build the address by hostname and connect to the server (host).
 - Convert can.cs.nthu.edu.tw/ into IP address.
 - Set the protocol. (Ex. use TCP protocol)
- Get the source code of a web page.
 - Send a HTTP GET request to the server. (GET xxx ... \r\n\r\n)
 - Receive responses from the server. (you may need to receive more than one responses)
- Show all hyperlinks the code found.
 - ``
 - If the string starts with "href" and the previous token ends with "a", then you find a hyperlink.

5. Submission

- a. Please provide a **markdown format pdf file (You can use hackmd and print it as a pdf file)** to show what functionalities your homework has.
 - For example, is it able to be compiled? Does it meet all requirements?
 - If you can run your program, please provide a screenshot to show how it works just like the examples in this document.
- b. Compress the source file(s) and related files (including readme.pdf) into **學號_作業_版本.zip (ex: 109062599_lab1_v1.zip)**
- c. Use **“mkdir ~/Lab1”** command to create a directory named “Lab1” at home directory and then upload the compressed file to the ~/Lab1 directory.
- d. Discussion is encouraged. However, **plagiarism is not allowed**. We will use, e.g., “Moss” for similarity comparison and 0 points will be given if plagiarism.
- e. You should **submit your assignment by the deadline**, or your assignment will not be graded, meaning that you will receive zero points.