



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN - ĐHQG TP.HCM
VNUHCM - UNIVERSITY OF SCIENCE

KHOA CÔNG NGHỆ THÔNG TIN

ĐỒ ÁN THỰC HÀNH MẠNG MÁY TÍNH:

LAB SOCKET

MÃ MÔN HỌC: CSC-10008_21CLC04

THỰC HIỆN: Nhóm 18, Lớp 21CLC04

GVHD: Đỗ Hoàng Cường

Nguyễn Thanh Quân

Huỳnh Thụy Bảo Trân

Tp. Hồ Chí Minh, tháng 11 năm 2022

MỤC LỤC

1.	<i>Thông tin của nhóm.....</i>	3
2.	<i>Đánh giá mức độ hoàn thành</i>	3
3.	<i>Kịch bản giao tiếp của chương trình Socket.....</i>	4
4.	<i>Môi trường lập trình</i>	6
5.	<i>Hướng dẫn sử dụng chương trình</i>	7
6.	<i>Bảng phân công công việc</i>	7
7.	<i>Các nguồn tài liệu tham khảo</i>	8

1. Thông tin của nhóm

Tên nhóm: 18

Mã số sinh viên	Họ và tên
21127627	Cao Nguyễn Khánh
21127711	Trịnh Minh Trung (Nhóm Trưởng)
21127535	Thành Thiện Nhân

2. Đánh giá mức độ hoàn thành

Chức năng	Ý nghĩa	Mức độ hoàn thành	Ghi chú
<i>Kết nối</i>	Cho phép kết nối đến server thông qua kết nối TCP	100%	
<i>Quản lý kết nối</i>	Khi client hoặc server ngắt kết nối đột ngột, không làm chương trình treo hay xảy ra lỗi	100%	
<i>Tải và lưu thành file dạng "Content-Length"</i>	Download file thành công	100%	
<i>Tải và lưu thành file dạng "Transfer-Encoding: chunked"</i>	Download file thành công	75%	Đối với một số trang, còn chưa xử lý được như google.com, bing.com, ...
<i>Tải các file trong folder</i>	Lưu tất cả các file trong folder	100%	

<i>Multiple Requests</i>	Gửi nhiều requests trong một connection khi download file trong folder	100%	
<i>Multiple Connection</i>	Concurrent, handle nhiều kết nối cùng lúc đến các web servers	100%	

3. Kịch bản giao tiếp của chương trình Socket

Trong chương trình Socket (Web Client) được nhóm 18 xây dựng, chính nó có chức năng như một Web Client giống như những Web Browser mà chúng ta hay sử dụng tuy nhiên ở đây chương trình chỉ cho phép người dùng tải về các file thuộc các loại như html, pdf, ppt, doc, ... mà được có trên web page thông qua giao thức HTTP (Hyper Text Transport Protocol) để truy xuất vào một máy chủ (Web Server) với một địa chỉ web (URL).

Khi đó, chương trình sẽ tạo một kết nối TCP/IP (thông qua một phần mềm socket) đến máy chủ (Host) của địa chỉ (URL) mà người dùng đã nhập với Port 80 của Web Server.

Sau khi kết nối thành công đến Web Server, Client sẽ gửi một yêu cầu (Request) với phương thức “GET” với giao thức HTTP phiên bản 1.1 (HTTP/1.1) kèm với tên của máy chủ (Host), loại điều khiển Socket là “keep-alive” (Connection: keep-alive) cùng với một số một số chỉ định về format, ngôn ngữ để thông báo cho Web Server rằng Client đang muốn truy xuất vào Web Page. Ví dụ về một yêu cầu theo phương thức “GET” cơ bản của giao thức HTTP:

```
GET /index.html HTTP/1.1\r\nHost:example.com\r\nConnection: keep-alive\r\n\r\n
```

Khi đó, Web Server sẽ nhận được một request, sau đó sẽ kiểm tra tên của Web Page trong request có tồn tại hay không hoặc bị lỗi. Dựa vào các trạng thái khác nhau đó, Web Server sẽ trả về các response khác nhau với các response code khác nhau như 200 cho một web page tồn tại, 404 cho một web page không được tìm thấy, ... sẽ được tìm

thấy ở dòng đầu trong một loạt các dòng Header Lines được phân cách bởi ký tự “\r\n” (“Accept-Ranges”, “Age”, “Cache-Control”, ...) dùng để cung cấp thông tin cho response mà Web Server trả về cho chương trình (Web Client).

Response của Web Server sẽ được chia ra làm hai phần được phân cách giữa ký tự “\r\n” gồm phần chứa các Header Lines và Body. Response này có kiểu dữ liệu ở dạng nhị phân.

```
HTTP/1.1 200 OK\r\n
Accept-Ranges: bytes\r\n
Age: 471725\r\n
Cache-Control: max-age=604800\r\n
...
Vary: Accept-Encoding\r\n
X-Cache: HIT\r\n
Content-Length: 1256\r\n
\r\n                                     Ký tự phân chia Headers và Body
<!doctype html>
<html>
<head>
  <title>Example Domain</title>
  ...
</body>
</html>
```

Ví dụ về một response của Web Server

Trong số các Header Lines của response từ Server sẽ chứa một header có thông tin để phân biệt loại của một Web Page như là Content Length kèm với tổng số byte (có kiểu dữ liệu là một số nguyên dạng thập phân) của phần thân của response (“Content-Length: <Số bytes>\r\n”), Transfer Encoding Chunked (“Transfer-Encoding: chunked”), ... Dựa vào Header Line này để có thể đọc được phần thân (Body) của một response vì phần body của mỗi dạng sẽ được hiển thị một cách khác nhau.

Đối với một Web Page trả về response có dạng Content Length, dữ liệu ở phần thân của response sẽ là một khối dữ liệu với kích thước bằng với số byte trong header line nói trên, mở đầu từ sau ký tự “\r\n” phân chia phần header và phần thân của một response và kéo dài cho đến hết response.

```
Content-Length: 1256\r\n
\r\n                                     Ký tự phân chia Headers và Body
<!doctype html>
<html>
<head>
  <title>Example Domain</title>
  ...
</body>
</html>
```

Ví dụ về response có dạng Content Length

Đối với một Web Page trả về response có dạng Transfer Encoding Chunked, phần thân của response sẽ được hiển thị bằng một chuỗi các khối dữ liệu còn được gọi là “chunk”. Mỗi chunk có kích thước dữ liệu nhất định, trước một chunk được hiển thị trong response là một số ở hệ thập lục phân dùng để biểu diễn kích thước của chunk đó dưới đơn vị byte, theo sau là một ký tự “\r\n” dùng để ngăn cách và nội dung của chunk đó với kích thước như trên. Phân biệt giữa hai chunk khác nhau là một ký tự điều khiển “\r\n”. Chunk cuối cùng sẽ rỗng và có kích thước là 0, theo sau đó là một ký tự điều khiển “\r\n”.

```
Transfer-Encoding: chunked
\r\n.                                     Ký tự phân chia Headers & Body
7\r\n
Mozilla\r\n
11\r\n
Developer Network\r\n
0\r\n
\r\n
```

Ví dụ về một response có dạng Transfer Encoding Chunked

4. Môi trường lập trình

- Môi trường lập trình: trên môi trường hệ điều hành cả Windows và MacOS
- Ngôn ngữ lập trình: Python
- Thư viện hỗ trợ lập trình socket: thư viện “socket” của Python
- Các thư viện được gọi: “re”, “os”, “pathlib”, “shutil”, “threading”, “time”
- Framework hỗ trợ để thực thi chương trình: không có

5. Hướng dẫn sử dụng chương trình

Sau khi chương trình được mở lên thành công, người dùng được yêu cầu nhập vào một hoặc nhiều (được ngăn cách bởi những dấu cách) đường link của Web Page (Web URL) mà người dùng muốn tải. Chương trình sẽ xem xét các URL vừa nhập mà xác định loại tải file mà người dùng muốn tải như tải file đơn hoặc là tải thư mục.

Có 3 cách để phân loại:

1. Với các đường link chỉ có tên của host thì mặc định tải và lưu tệp index.html. Tệp tải về được lưu vào đường dẫn thư mục cùng cấp với thư mục chứa chương trình.
2. Với các đường link con có file extension rõ ràng thì tệp được lưu về dưới dạng của extension đó như *.docx, *.ppt, *.pdf, ... Tệp tải về được lưu vào đường dẫn thư mục cùng cấp với thư mục chứa chương trình.
3. Với các đường link có thư mục chứa các file con thì tải các tệp có chứa trong thư mục đó. Các tệp con được tải vào một thư mục được tạo mới nằm cùng cấp với đường dẫn thư mục chứa chương trình. Quá trình tải về các tệp trong thư mục này được thực hiện cùng một lúc, chứ không tuần tự.

6. Bảng phân công công việc

Công việc	Chi tiết công việc	Người thực hiện
Kết nối	Kết nối với server bằng thư viện “socket” của python	Cả nhóm
Xử lý URL Website	Phân tích URL người dùng nhập vào để lấy domain và tên tệp để phục vụ cho việc tải tệp và xác định loại tải xuống là tệp đơn hay là thư mục	Thành Thiện Nhân
Gửi yêu cầu cho Server	Tìm hiểu về cấu trúc của một request HTTP và gửi đến Server	Thành Thiện Nhân
Quản lý kết nối	Khi Client hoặc Server mất kết nối đột ngột, thử kết nối lại	Thành Thiện Nhân

Xác định loại tệp cần tải	Đọc vào phần Header Lines của response và cho biết tệp cần tải ở dạng Content Length hay Transfer Encoding Chunked	<i>Cao Nguyễn Khánh</i>
Đọc nội dung response dạng Content Length	Đọc nội dung của phần thân response dựa vào số byte content ở trên Header	<i>Cao Nguyễn Khánh</i>
Đọc nội dung response dạng Transfer Encoding Chunked	Đọc nội dung của phần thân response bằng cách đọc từng chunk dữ liệu dựa vào độ lớn mỗi chunk	<i>Trịnh Minh Trung</i>
Tải xuống một tệp	Tải một tệp bất kì về đường dẫn thư mục cùng cấp với chương trình Client bằng nội dung nhị phân đã đọc được	<i>Cả nhóm</i>
Lập trình đa luồng cho chương trình	Áp dụng đa luồng vào chương trình để hỗ trợ tải thư mục và cho phép người dùng nhập nhiều URL ở đầu chương trình	<i>Trịnh Minh Trung</i>
Viết báo cáo	Báo cáo về kịch bản giao tiếp của chương trình	<i>Trịnh Minh Trung</i>

7. Các nguồn tài liệu tham khảo

Sau đây là các đường link của các trang web mà nhóm đã dùng để tìm hiểu về ngôn ngữ lập trình python, lập trình socket và lập trình socket thông qua ngôn ngữ python:

- internalpointers.com
- stackoverflow.com
- docs.python.org
- w3schools.com
- superfastpython.com
- quickprogrammingtips.com
- codethief.io
- geeksforgeeks.com
- simplilearn.com
- realpython.com
- net-informations.com

- instructobit.com
- flexiple.com
- thispointer.com
- digitalocean.com
- blog.finxter.com
- developer.mozilla.org