

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN**



BÁO CÁO

ĐỒ ÁN THỰC HÀNH – LẬP TRÌNH SOCKET

MÔN MẠNG MÁY TÍNH

NHÓM THỰC HIỆN:

21120576 – Trần Đình Nhật Trí
21120575 – Nguyễn Thành Trí
21120588 – Nguyễn Phước Anh Tuấn

Giảng viên lý thuyết: Nguyễn Thanh Quân

Lớp: 21CTT5

Học kỳ: HK1/ 2022-2023

MỤC LỤC

I. THÔNG TIN THÀNH VIÊN	3
II. ĐÁNH GIÁ MỨC ĐỘ HOÀN THÀNH	3
III. BẢNG PHÂN CÔNG CÔNG VIỆC TRONG ĐỒ ÁN.....	3
IV. KỊCH BẢN GIAO TIẾP CHƯƠNG TRÌNH.....	5
1. <i>Giao thức trao đổi giữa Client và Server.....</i>	<i>5</i>
2. <i>Cấu trúc thông điệp.....</i>	<i>5</i>
3. <i>Kiểu dữ liệu thông điệp</i>	<i>5</i>
V. MÔI TRƯỜNG LẬP TRÌNH VÀ CÁC FRAMEWORK HỖ TRỢ ĐỂ THỰC THI ỨNG DỤNG	5
VI. HƯỚNG DẪN SỬ DỤNG CÁC TÍNH NĂNG CHƯƠNG TRÌNH.....	7
VII. TÀI LIỆU THAM KHẢO.....	8

THÔNG TIN CHUNG VỀ ĐỒ ÁN

I. THÔNG TIN THÀNH VIÊN

MÃ SỐ SINH VIÊN	HỌ VÀ TÊN
21120576	Trần Đình Nhật Trí
21120575	Nguyễn Thành Trí
21120588	Nguyễn Phước Anh Tuấn

II. ĐÁNH GIÁ MỨC ĐỘ HOÀN THÀNH

CHỨC NĂNG	MỨC ĐỘ HOÀN THÀNH	LỖI
KẾT NỐI	100%	
QUẢN LÝ KẾT NỐI	100%	
TẢI VÀ LƯU THÀNH FILE DẠNG "CONTENT-LENGTH"	100%	
TẢI VÀ LƯU THÀNH FILE DẠNG "TRANSFER-ENCODING: CHUNKED"	100%	
TẢI CÁC FILE TRONG FOLDER	90%	Chưa thể check connection: closed trong header
MULTIPLE REQUESTS	100%	
MULTIPLE CONNECTIONS	100%	

III. BẢNG PHÂN CÔNG CÔNG VIỆC TRONG ĐỒ ÁN

Bảng phân công công việc trong đồ án:

CHỨC NĂNG	NGƯỜI THỰC HIỆN
KẾT NỐI	Nguyễn Thành Trí

QUẢN LÝ KẾT NỐI	Trần Đình Nhật Trí
TẢI VÀ LƯU THÀNH FILE DẠNG "CONTENT-LENGTH"	Nguyễn Phước Anh Tuấn
TẢI VÀ LƯU THÀNH FILE DẠNG "<i>TRANSFER-ENCODING: CHUNKED</i>"	Nguyễn Thành Trí
TẢI CÁC FILE TRONG FOLDER	Trần Đình Nhật Trí
MULTIPLE REQUESTS	Nguyễn Thành Trí
MULTIPLE CONNECTIONS	Nguyễn Phước Anh Tuấn
REPORT	Trần Đình Nhật Trí

NỘI DUNG ĐỒ ÁN

IV. KỊCH BẢN GIAO TIẾP CHƯƠNG TRÌNH

1. Giao thức trao đổi giữa Client và Server

- Giao thức trao đổi được sử dụng là giao thức TCP/IP ở tầng Transport.

2. Cấu trúc thông điệp

- Đầu tiên, để các Client có thể gửi nhận dữ liệu thì ta phải tạo socket và kết nối client tới server bằng domain name được nhập vào từ người dùng với port là 80. Nếu tạo và kết nối thành công, Client sẽ bắt đầu gửi request tới server. Nếu tạo socket hoặc kết nối không thành công thì in ra lỗi và đóng chương trình.
- Client sẽ gửi request cho server, server sẽ response và client sẽ xử lý theo đúng định dạng response mà server gửi về.
- Khi tải xong hoặc client đóng kết nối thì sẽ gửi thông điệp đóng kết nối tới server hoặc nếu client mất kết nối đột ngột thì sẽ in ra lỗi ngắt kết nối và đóng chương trình.

3. Kiểu dữ liệu thông điệp

- Dữ liệu nhận vào và gửi đi đều thuộc kiểu dữ liệu String (chuỗi ký tự).
- Dữ liệu được mã hoá và giải mã kiểu UTF-8.

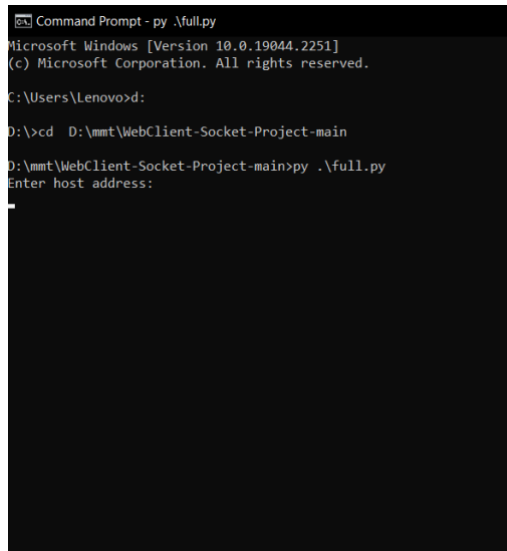
V. MÔI TRƯỜNG LẬP TRÌNH VÀ CÁC FRAMEWORK HỖ TRỢ ĐỂ THỰC THI ỨNG DỤNG

- Ứng dụng được viết và chạy trên hệ điều hành MacOS, Windows 10 và 11.
- Ngôn ngữ lập trình: Python (Phiên bản 3.10.8 trở lên).
- Hệ điều hành: MacOS, Windows.
- Các thư viện hỗ trợ dùng để viết chương trình:
 - socket: thực hiện các thao tác socket.
 - urllib.parse: xử lý url truyền vào.
 - re: tìm và xử lý biểu thức chính quy.
 - os: xử lý thao tác tạo thư mục.
 - concurrent.futures: xử lý đa kết nối tới nhiều server.
 - from os import path: lấy basename của url để đặt tên file.
 - sys: thực hiện lệnh đóng chương trình khi có lỗi.

- Nếu máy tính chưa cài đặt thư viện thì có thể sử dụng lệnh
 - Windows: `pip install <thư viện>` trong cmd (nếu như import thư viện bị lỗi)
 - MacOS: `pip3 install <thư viện>` trong terminal.
 - Tải pip:
 - Windows: `python get-pip.py`
 - MacOS: `curl https://bootstrap.pypa.io/get-pip.py -o get-pip.py, python3 get-pip.py`

VI. HƯỚNG DẪN SỬ DỤNG CÁC TÍNH NĂNG CHƯƠNG TRÌNH

1. Mở terminal và nhập với cú pháp: `py .\tên file.py <url 1><space><url 2>` (<> là các tham số có thể truyền hoặc không)
 - **TH1:** Nếu người dùng không nhập thêm đường dẫn nào khác thì sẽ bắt đầu chạy chương trình python với url được nhập thêm như dưới:



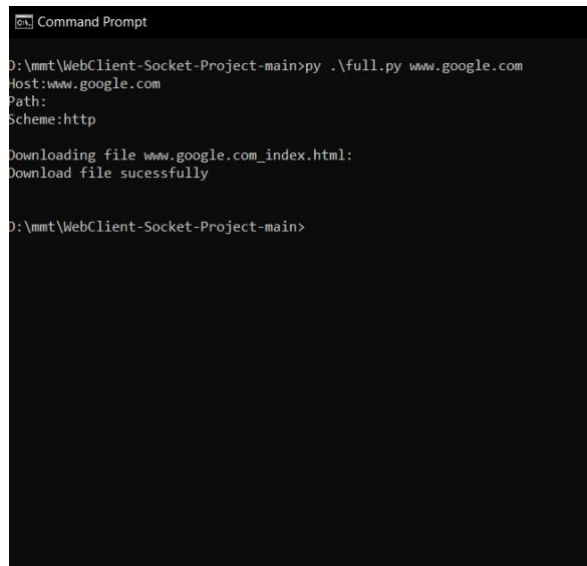
```
Command Prompt - py .\full.py
Microsoft Windows [Version 10.0.19044.2251]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Lenovo>cd D:\mnt\WebClient-socket-project-main

D:\mnt\WebClient-socket-project-main>py .\full.py
Enter host address:
```

Hình 1

- **TH2:** Nếu người dùng nhập thêm đường dẫn thì sẽ bắt đầu chạy chương trình trực tiếp dựa vào số url mà người dùng nhập. Mỗi url phải được cách bởi 1 ký tự space.



```
Command Prompt

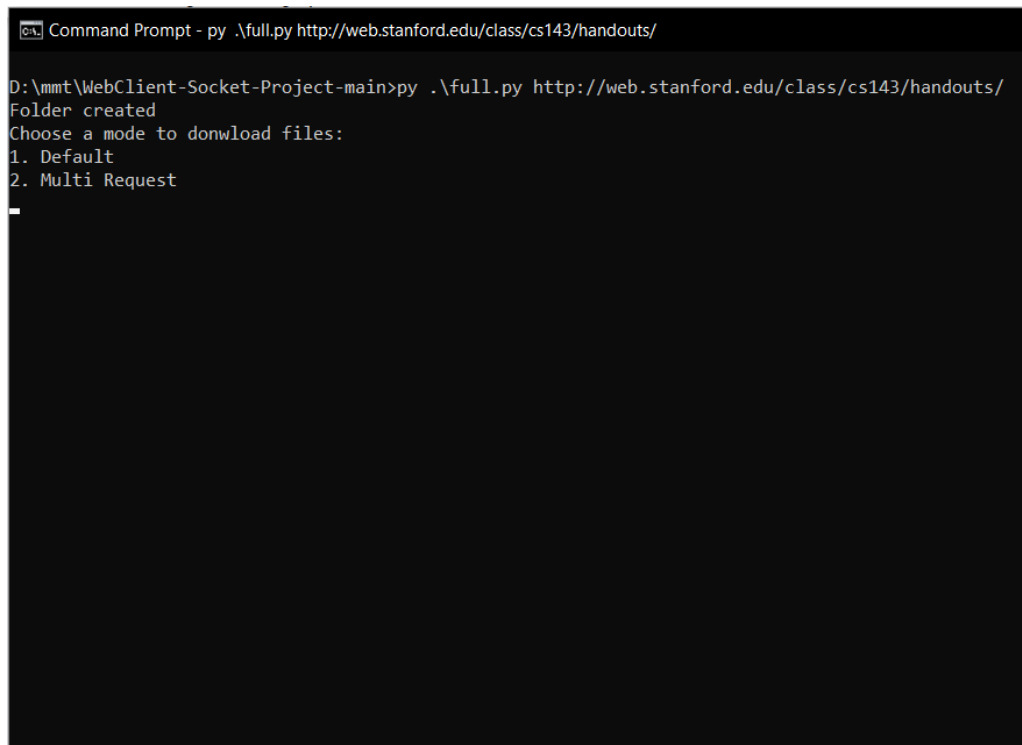
D:\mnt\WebClient-socket-project-main>py .\full.py www.google.com
Host:www.google.com
Path:
Scheme:http

Downloading file www.google.com_index.html:
Download file sucessfully

D:\mnt\WebClient-socket-project-main>
```

Hình 2

2. Với trường hợp khi người dùng truyền vào 1 đường dẫn để tải folder, người dùng sẽ có 2 hình thức kết nối là request đơn cho mỗi file được tải, và đa request cho một lần tải.



```
Command Prompt - py .\full.py http://web.stanford.edu/class/cs143/handouts/

D:\mmt\WebClient-Socket-Project-main>py .\full.py http://web.stanford.edu/class/cs143/handouts/
Folder created
Choose a mode to download files:
1. Default
2. Multi Request
-
```

Hình 3

VII. TÀI LIỆU THAM KHẢO

Python: <https://www.python.org/doc/>

Socket: <https://docs.python.org/3/library/socket.html>

<https://realpython.com/python-sockets/>

<https://www.scrapingbee.com/blog/web-scraping-101-with-python/>

Parallel connections: <https://www.oreilly.com/library/view/http-the-definitive/1565925092/ch04s04.html>

Multiple connections:

<https://superfastpython.com/threadpoolexecutor-download-files/>

<https://docs.python.org/3/library/concurrent.futures.html#concurrent.futures.Future>

RFC7230: <https://datatracker.ietf.org/doc/html/rfc7230#section-3>

Exception handling: https://www.youtube.com/watch?v=NIWwJbo-9_8

https://www.w3schools.com/python/python_try_except.asp

Các tài liệu thực hành được đăng trên Moodle môn Mạng máy tính 21CTT5 – Khoa Công nghệ Thông tin – Trường Đại học Khoa học Tự nhiên – ĐHQG TPHCM.