

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
SOCKET

Môn học: Mạng máy tính

♣GIÁO VIÊN HƯỚNG DẪN♣

Giáo viên lý thuyết:

Huỳnh Thụy Bảo Trân

Giáo viên thực hành:

Chung Thùy Linh

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
SOCKET

Môn học: Mạng máy tính

Đề 2: Web Client

♣ **THÀNH VIÊN NHÓM** ♣

21120313 – Trần Nam Phương

21120317 – Nguyễn Phan Anh Quốc

21120321 – Nguyễn Văn Siêu

MỤC LỤC

MỤC LỤC.....	3
THÔNG TIN THÀNH VIÊN.....	4
ĐÁNH GIÁ MỨC ĐỘ HOÀN THÀNH	4
KỊCH BẢN GIAO TIẾP CỦA CHƯƠNG TRÌNH.....	5
1. Giao thức trao đổi:	5
2. Cấu trúc thông điệp:	5
3. Kiểu dữ liệu thông điệp:	5
MÔI TRƯỜNG LẬP TRÌNH.....	6
HƯỚNG DẪN SỬ DỤNG CHƯƠNG TRÌNH.....	7
BẢNG PHÂN CÔNG CÔNG VIỆC.....	9
NGUỒN THAM KHẢO.....	10

THÔNG TIN THÀNH VIÊN

Mã số sinh viên	Họ và tên	Chú thích
21120313	Trần Nam Phương	21120313@student.hcmus.edu.vn
21120317	Nguyễn Phan Anh Quốc	21120317@student.hcmus.edu.vn
21120321	Nguyễn Văn Siêu	21120321@student.hcmus.edu.vn

ĐÁNH GIÁ MỨC ĐỘ HOÀN THÀNH

STT	Chức năng	Đánh giá	Chú thích
1	Kết Nối	100%	
2	Quản lý kết nối	100%	
3	Tải và lưu thành file dạng “Content-length”	100%	
4	Tải và lưu thành file dạng “Transfer-encoding: chunked”	50%	Chưa xử lý được việc web google trả về chunk size lớn nên lưu dữ liệu còn thiếu
5	Tải các file trong folder	100%	
6	Multiple request	0%	Chưa xử lý được việc server đóng kết nối sau khi trả về response thứ nhất
7	Multiple connection	100%	

KỊCH BẢN GIAO TIẾP CỦA CHƯƠNG TRÌNH

1. Giao thức trao đổi:

Giao thức trao đổi giữa Client và Server được sử dụng trong đồ án là TCP.

2. Cấu trúc thông điệp:

Khi Client muốn truy suất thông tin của Server thì Client sẽ gửi cho Sever thông điệp có cấu trúc như sau:

GET **resource** HTTP/1.1\r\n

Host: **host**\r\n

Connection: Keep-Alive\r\n

\r\n

Mỗi dòng sẽ kết thúc với '\r\n' và toàn bộ thông điệp cũng sẽ kết thúc với '\r\n'.

Trong đó, **resource** là tài nguyên mà client cần truy cập tới trong server, **host** là trang chủ của server đó. Sau khi nhận được thông điệp từ client, server sẽ phản hồi dữ liệu lại cho client.

3. Kiểu dữ liệu thông điệp:

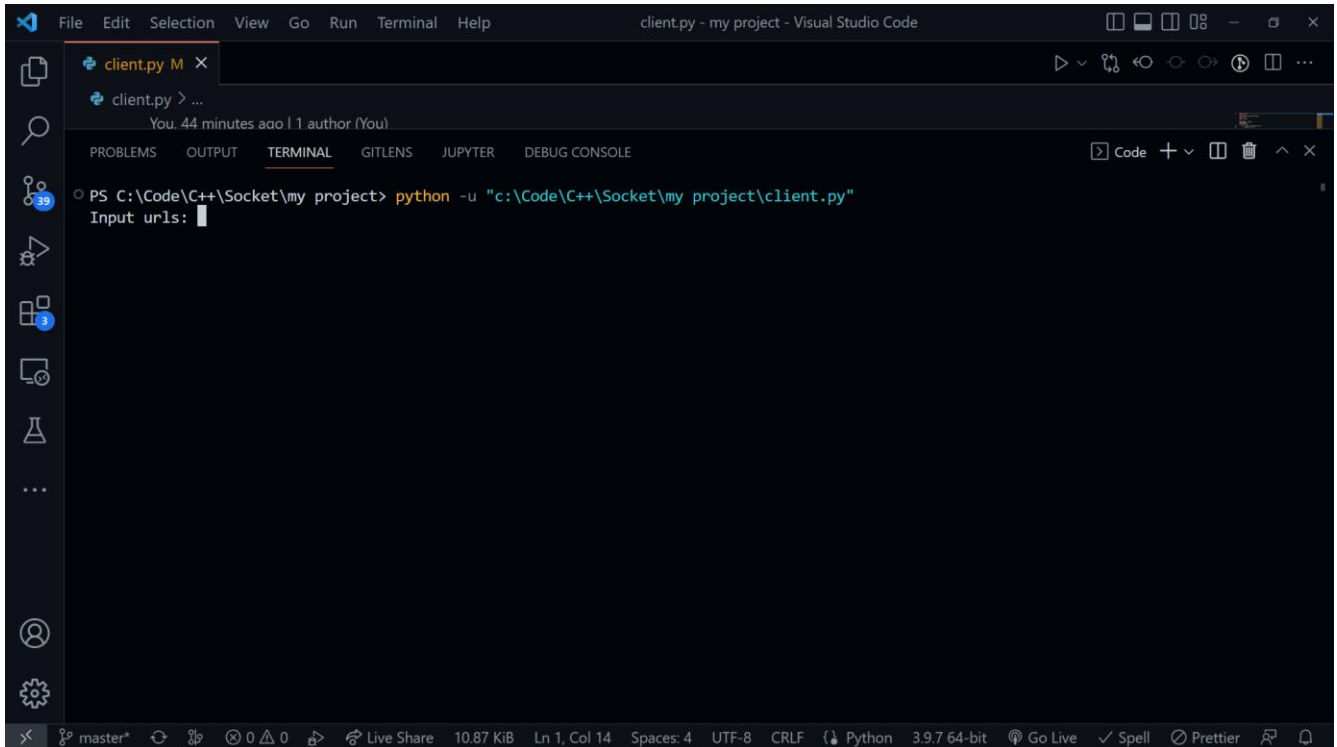
Dữ liệu được gửi đến và dữ liệu được nhận vào đều là kiểu chuỗi ký tự (string).

MÔI TRƯỜNG LẬP TRÌNH

- Ứng dụng viết trên hệ điều hành **Windows 11**.
- Ngôn ngữ lập trình **Python** (Phiên bản 3.9.7 64-bit).
- Các thư viện hỗ trợ: **socket, threading, re, os, time**.

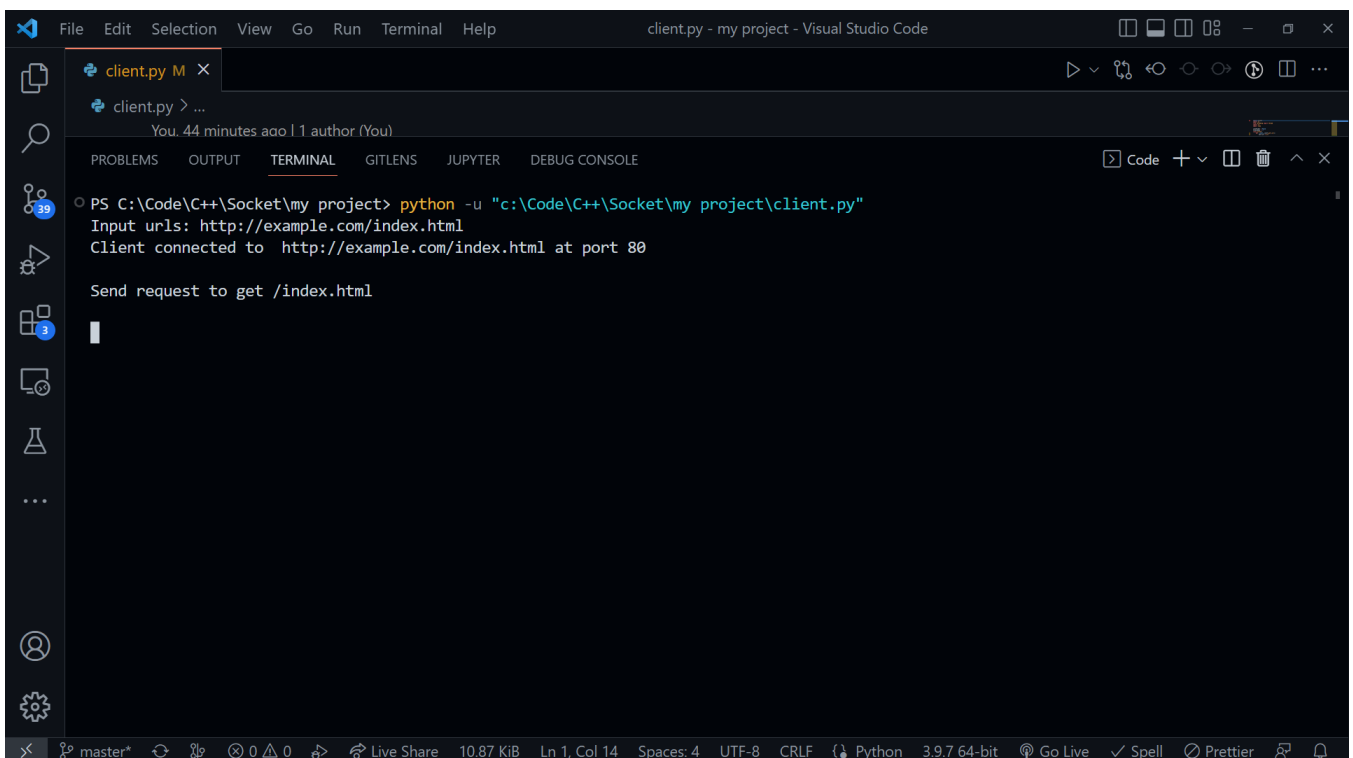
HƯỚNG DẪN SỬ DỤNG CHƯƠNG TRÌNH

Khi bắt đầu chạy file client.py thì chương trình sẽ yêu cầu nhập vào link url của các trang web cần tải.



```
client.py - my project - Visual Studio Code
client.py M x
client.py > ...
You. 44 minutes ago | 1 author (You)
PROBLEMS OUTPUT TERMINAL GITLENS JUPYTER DEBUG CONSOLE
PS C:\Code\C++\Socket\my project> python -u "c:\Code\C++\Socket\my project\client.py"
Input url:
```

Sau khi nhập thì Client tiến hành kết nối tới server và gửi request.



```
client.py - my project - Visual Studio Code
client.py M x
client.py > ...
You. 44 minutes ago | 1 author (You)
PROBLEMS OUTPUT TERMINAL GITLENS JUPYTER DEBUG CONSOLE
PS C:\Code\C++\Socket\my project> python -u "c:\Code\C++\Socket\my project\client.py"
Input url: http://example.com/index.html
Client connected to http://example.com/index.html at port 80
Send request to get /index.html
```

Sau khi tải xong sẽ in địa chỉ của file đã tải .

The screenshot shows a Visual Studio Code window with a terminal open. The terminal output shows a successful HTTP GET request using curl. The command executed is `curl -u 'http://example.com/index.html' http://example.com/index.html`. The output indicates that the client connected to the URL at port 80 and successfully retrieved the content, which was saved in `C:\Code\C++\Socket\my project\index.html`.

```
PS C:\Code\C++\Socket\my project> curl -u 'http://example.com/index.html' http://example.com/index.html
Client connected to http://example.com/index.html at port 80

Send request to get /index.html

Saved in C:\Code\C++\Socket\my project\index.html

PS C:\Code\C++\Socket\my project>
```

Nếu nhập nhiều url thì các url cách nhau bằng khoảng trắng

The screenshot shows the Visual Studio Code editor with a terminal window open. The terminal is running a command prompt where the user has entered the command `python -u "c:\Code\C++\Socket\my project\client.py"`. The output of the command is `Input urls: http://web.stanford.edu/class/cs231a/project.html http://web.stanford.edu/class/cs224w/slides/08-GNN-application.pdf http://example.com/index.html`. The interface includes a sidebar on the left with icons for Explorer, Search, Source Control, Run and Debug, Extensions, and Settings. The top bar shows the file name `client.py` and the project name `my project`. The bottom status bar displays information about the current file, including the branch `master*`, the file size `10.87 KiB`, the current line and column `Ln 2, Col 10`, the number of spaces `Spaces: 4`, the encoding `UTF-8`, the line ending `CRLF`, the Python version `Python 3.9.7 64-bit`, and the status of various extensions like `Go Live`, `Spell`, `Prettier`, and `Live Share`.

BẢNG PHÂN CÔNG CÔNG VIỆC

Họ và tên – MSSV	Công việc
Trần Nam Phương 21120313	Gửi request tới server
	Xử lý header response
	Xử lý body response và lưu về dạng “Content-length”
	Báo cáo đồ án
Nguyễn Phan Anh Quốc 21120317	Tạo và quản lý kết nối đến server
	Phân tích url để lấy địa chỉ host và tên file tải về
	Thực hiện multiple request
	Xử lý đa luồng để gửi nhiều kết nối đến server(multiple connection)
Nguyễn Văn Siêu 21120321	Xử lý body response và lưu về dạng “Transfer-encoding: chunked”
	Xử lý dữ liệu html để lấy url của các file cần download trong folder
	Xử lý đa luồng để tải tất cả file trong folder

NGUỒN THAM KHẢO

- Đa luồng – Thread
https://www.geeksforgeeks.org/multithreading-python-set-1/?fbclid=IwAR0iYlwmMhxbbM7gA-LAoMKMygwQOES6m9PIJh9UYU5nee_vv0nbvuDIRHM
- HTTP Header Response
<https://developer.mozilla.org/en-US/docs/Web/HTTP/Headers>
- Write file in Python and save in computer
https://www.w3schools.com/python/python_file_write.asp
<https://www.geeksforgeeks.org/python-os-path-join-method/>
- Transfer-encoding: chunked
<https://bunny.net/academy/http/what-is-chunked-encoding/>
- Thư viện re
https://www.w3schools.com/python/python_regex.asp?fbclid=IwAR2RgTZ8mOm77SIUAJay6EmXaHvzpWv0ImwIQr8oeWHPPBU8ZVnzD0qCgXk
https://docs.python.org/3/library/re.html?fbclid=IwAR2q9A9bZzE4W5XHAbzrjvtc8ewHITqzDqt_Pre9o-2Sym495odn24Sdu64
- Và các tài liệu hướng dẫn thực hành môn học đã cung cấp.