



University of Science - VNUHCM

21CTT5 - fit@hcmus

The report of
SOCKET PRACTICE - WEB CLIENT

Giảng viên

Nguyễn Thanh Quân

Thành viên nhóm

21120546 - Nguyễn Thanh Sang

21120553 - Trần Thái Tân

21120566 - Nguyễn Hữu Thuận

<i>I. Thông tin nhóm</i>	<i>2</i>
<i>II. Đánh giá mức độ hoàn thành</i>	<i>2</i>
<i>III. Kịch bản giao tiếp của chương trình</i>	<i>2</i>
1. Giao thức trao đổi giữa client và server	2
2. Cấu trúc và kiểu dữ liệu của message	2
<i>IV. Môi trường lập trình và các framework hỗ trợ</i>	<i>3</i>
<i>V. Hướng dẫn sử dụng</i>	<i>3</i>
<i>VI. Bảng phân công công việc</i>	<i>5</i>
<i>VII. Tài liệu tham khảo</i>	<i>5</i>

I. Thông tin nhóm

21120546 - Nguyễn Thanh Sang

21120553 - Trần Thái Tân

21120566 - Nguyễn Hữu Thuận

II. Đánh giá mức độ hoàn thành

Đã hoàn thành 8/8 mục, cụ thể:

1	Kết nối	(0,5 điểm)	100%
	Cho phép client kết nối đến server thông qua kết nối TCP		
2	Quản lí kết nối	(0,5 điểm)	100%
	Khi client hoặc server mất kết nối đột ngột, không làm chương trình treo hay xảy ra lỗi		
3	Tải và lưu file dạng “Content-Length”	(3,0 điểm)	100%
	Download file thành công		
4	Tải và lưu thành file dạng "Transfer-Encoding: chunked"	(1,5 điểm)	100%
	Download file thành công		
5	Tải các file trong folder	(1,5 điểm)	100%
	Lưu tất cả các file trong folder		
6	*Multiple requests	(1,0 điểm)	100%
	Gửi nhiều requests trong một connection khi download file trong folder		
7	*Multiple connection	(1,0 điểm)	100%
	Concurrent, handle nhiều kết nối cùng lúc đến các web servers		
8	Report	(1,0 điểm)	100%
	Download file thành công		

III. Kịch bản giao tiếp của chương trình

1. Giao thức trao đổi giữa client và server

Giao thức TCP/IP.

2. Cấu trúc và kiểu dữ liệu của message

2.1. Message gửi từ client đến server

- Để truy cập đến một trang web với URL cho trước, phía client sẽ tạo 1 kết nối đến server với port 80.
- Sau khi kết nối đến web server, client sẽ gửi request đến server để tải page (với cấu trúc request như trong tài liệu bài giảng).

- Client sẽ tính toán dựa vào phần header nhận được để tải page trong trường hợp “Content-Length” hoặc “Transfer-Encoding: chunked”.

2.2. Message gửi từ server đến client

- Client sẽ nhận từ server với dữ liệu là từng byte một và client sẽ chỉ tạo ghi file page nhận được mà không ghi phần header.
- Client sẽ kiểm tra trong header HTTP status code để viết và báo ra những lệnh thích hợp.
- Client sẽ trả về những file/folder cần thiết như yêu cầu đề bài.
- Tùy vào yêu cầu, client sẽ sử dụng những phương thức khác nhau để tải file, ví dụ tải bằng một socket với nhiều file trong một folder, tải mỗi file bằng một socket, tải nhiều file cùng lúc với cơ chế đa luồng,... (phần này được trình bày bằng code)

IV. Môi trường lập trình và các framework hỗ trợ

- Môi trường lập trình: Visual Studio Code, Emacs Doom.
- Ngôn ngữ lập trình: Java.
- Thư viện hỗ trợ: java.io, java.net.socket, java.net.SocketTimeoutException, java.util.ArrayList.

V. Hướng dẫn sử dụng

- Mở thư mục **src** với **terminal**.

```
C:\Users\ADMIN\OneDrive - VNU-HCMUS\Desktop\Web-Socket\src>
```

- Các chế độ download:

- + Tải một file hoặc một folder:

java -jar Main.jar [URL]

- Nếu tải folder thì mặc định là tải với nhiều socket (mỗi lần tải một file là tạo và sử dụng một socket mới).
- Ví dụ:

java -jar Main.jar http://web.stanford.edu/class/cs231a/assignments.html

```
C:\Users\ADMIN\OneDrive - VNU-HCMUS\Desktop\Web-Socket\src>java -jar Main.jar http://web.stanford.edu/class/cs231a/assignments.html
Download 1 url
Download file
Downloading assignments.html...
Downloaded assignments.html!
Closed socket
```

- + Tải một folder theo dạng **multiple requests** (dùng socket ban đầu để tải các file trong folder):

java -jar Main.jar [URL] -s

- Ví dụ:

```
C:\Users\ADMIN\OneDrive - VNU-HCMUS\Desktop\Web-Socket\src>java -jar Main.jar http://web.stanford.edu/class/cs143/handouts/ -s
Download folder with single sockets
Downloading PA1%20Handout.pdf...
Downloaded PA1%20Handout.pdf!
File web.stanford.edu/class/cs143/handouts/PA1%20Handout.pdf downloaded!
Downloading PA2%20Handout.pdf...
Downloaded PA2%20Handout.pdf!
File web.stanford.edu/class/cs143/handouts/PA2%20Handout.pdf downloaded!
Downloading PA3%20Handout.pdf...
Downloaded PA3%20Handout.pdf!
File web.stanford.edu/class/cs143/handouts/PA3%20Handout.pdf downloaded!
Downloading PA4%20Handout.pdf...
Downloaded PA4%20Handout.pdf!
File web.stanford.edu/class/cs143/handouts/PA4%20Handout.pdf downloaded!
Downloading PA5%20Handout.pdf...
Downloaded PA5%20Handout.pdf!
File web.stanford.edu/class/cs143/handouts/PA5%20Handout.pdf downloaded!
Downloading WA1.pdf...
Downloaded WA1.pdf!
File web.stanford.edu/class/cs143/handouts/WA1.pdf downloaded!
Downloading WA1_sol.pdf...
Downloaded WA1_sol.pdf!
File web.stanford.edu/class/cs143/handouts/WA1_sol.pdf downloaded!
Downloading WA1_template.tex...
Downloaded WA1_template.tex!
File web.stanford.edu/class/cs143/handouts/WA1_template.tex downloaded!
Downloading WA2.pdf...
Downloaded WA2.pdf!
File web.stanford.edu/class/cs143/handouts/WA2.pdf downloaded!
Downloading WA2_sol.pdf...
Downloaded WA2_sol.pdf!
File web.stanford.edu/class/cs143/handouts/WA2_sol.pdf downloaded!
Downloading WA2_template.tex...
```

- + Tải nhiều URL cùng lúc (sử dụng cơ chế đa luồng trong Java):

java -jar Main.jar [URL1] [URL2] ...

- Ví dụ:

```
C:\Users\ADMIN\OneDrive - VNU-HCMUS\Desktop\Web-Socket\src>java -jar Main.jar http://www-net.cs.umass.edu/wireshark-labs/Wireshark_Intro_v8.1.docx http://www.httpwatch.com/httpgallery/chunked/chunkedimage.aspx
Download multithread with multiple URLs
www-net.cs.umass.edu/wireshark-labs/Wireshark_Intro_v8.1.docx
Downloading Wireshark_Intro_v8.1.docx...
www.httpwatch.com/httpgallery/chunked/chunkedimage.aspx
Downloading chunkedimage.aspx...
Downloaded Wireshark_Intro_v8.1.docx!
Closed socket
Downloaded chunkedimage.aspx!
Closed socket
```

- + Nếu không nhập URL, chương trình sẽ thông báo ra ra yêu cầu nhập URL:

```
C:\Users\ADMIN\OneDrive - VNU-HCMUS\Desktop\Web-Socket\src>java -jar Main.jar
Please enter a URL
```

- Socket sẽ đóng hoặc xuất ra dòng lệnh tương ứng nếu thời gian mạng bị ngắt kết nối quá so với thời gian đã set timeout trước đó:

```
C:\Users\ADMIN\OneDrive - VNU-HCMUS\Desktop\Web-Socket\src>java -jar Main.jar http://web.stanford.edu/class/cs143/handouts/ -s
Download folder with single sockets
Downloading PA1%20Handout.pdf...
Downloaded PA1%20Handout.pdf!
File web.stanford.edu/class/cs143/handouts/PA1%20Handout.pdf downloaded!
Downloading PA2%20Handout.pdf...
Connection closed
```

(File PA2%20Handout.pdf đang tải thì bị mất kết nối, nó sẽ đợi một khoảng thời gian và nếu qua khoảng thời gian đó thì socket sẽ đóng và thoát chương trình)

Lưu ý: yêu cầu người sử dụng phải cài đặt môi trường java version từ 17 trở lên.

VI. Bảng phân công công việc

Phần	Nguyễn Thanh Sang	Trần Thái Tân	Nguyễn Hữu Thuận	Phần trăm
Kết nối	0%	0%	5%	5%
Quản lí kết nối	0%	5%	0%	5%
Content-Length	30%	0%	0%	30%
Transfer-Encoding: chunked	0%	15%	0%	15%
Folder	0%	5%	10%	15%
Multiple requests	0%	10%	0%	10%
Multiple connection	0%	0%	10%	10%
Report	2%	0%	8%	10%
Tổng cộng	32%	35%	33%	100%

VII. Tài liệu tham khảo

1. Seminar socket - thầy Nguyễn Thanh Quân (<https://drive.google.com/drive/folders/1v0tqARrWoFuk9IJ1tZ4-FNTE4h4yKhpt?usp=sharing>)
2. Slide bài giảng tiếng Việt - HCMUS (<https://drive.google.com/drive/folders/0BwRvtIYjNsvrN00yMnMtQ2dVeVE?resourcekey=0-G9yiQm41VkGHPNYrR4PGgw>)
3. Slide bài giảng tiếng Anh (<https://drive.google.com/drive/folders/14GxuAljtpMsjkX2rmzHr-xnTFcFzNvft>)
4. Video lập trình mạng Java (<https://www.youtube.com/watch?v=afab2rUkWRQ&t=1356s>)
5. <https://zetcode.com/java/socket/>
6. <https://stackoverflow.com/questions/33373334/java-http-server-sending-chunked-response>
7. <https://stackoverflow.com/questions/487632/java-http-download-corrupts-file>
8. <https://stackoverflow.com/questions/36471065/why-is-faster-to-open-a-new-socket-for-each-communication-that-sharing-same-sock>
9. <https://stackoverflow.com/questions/37100100/i-want-to-test-java-multithread-but-only-one-thread-run>
10. <https://stackoverflow.com/questions/28148060/multi-threaded-java-tcp-client>