

### Περιγραφή άσκησης 3

Η άσκηση αυτή αποτελείται από 3 τμήματα, εκ των οποίων το πρώτο είναι γραμμένο σε κέλυφος bash και τα άλλα δύο σε C.

Το πρώτο μέρος, ο WebCreator, δημιουργεί  $w \cdot p$  σελίδες html, σύμφωνα με τον αλγόριθμο που έχει δοθεί στην εκφώνηση της άσκησης. Κάθε τέτοια σελίδα έχει ως source file το βιβλίο που προτείνεται σαν link στην εκφώνηση, ενώ αποτελείται από  $m$  γραμμές και  $f+g$  συνδέσμους. Αν ο χρήστης δώσει είσοδο  $w=2$  και  $p=2$ , τότε είναι σίγουρο ότι το αρχείο όλες οι σελίδες θα έχουν τουλάχιστον έναν εισερχόμενο σύνδεσμο. Η επαλήθευση αυτή γίνεται με τη σύγκριση των unique ονομάτων όλων των συνδέσμων που υπάρχουν στα αρχεία με το σύνολο των συνδέσμων που δημιουργήθηκαν τυχαία. Αν τα δύο αυτά σύνολα είναι ίδια, τότε υπάρχει ένας τουλάχιστον εισερχόμενος σύνδεσμος για κάθε σελίδα, αλλιώς όχι.

Το δεύτερο μέρος, ο WebServer, αποτελείται από 4 ομάδες νημάτων. Το ένα νήμα είναι η main που δημιουργεί όλα τα ακόλουθα νήματα. Το δεύτερο νήμα είναι ο producer, ο οποίος «υποδέχεται» κάθε πελάτη στην πόρτα -p και τοποθετεί το socket του πελάτη σε μια κοινόχρηστη ουρά. Στην ουρά αυτή έχουν πρόσβαση οι consumers (τρίτη ομάδα νημάτων), οι οποίοι πρακτικά εξυπηρετούν τον εκάστοτε πελάτη και τερματίζουν αμέσως μετά. Τέλος, υπάρχει και το νήμα commander που εξυπηρετεί πελάτες (που έρχονται κυρίως μέσω telnet) στην πόρτα -c. Στην περίπτωση SHUTDOWN, στέλνεται από τον commander ειδικό μήνυμα τερματισμού στους consumers, ενώ ο producer τερματίζεται με pthread\_cancel.

Ο WebCrawler αποτελείται από 3 ομάδες νημάτων, δηλαδή τη main, τον producer και τους consumers. Αυτή τη φορά, ο producer τοποθετεί το αρχικό url στην ουρά και στη συνέχεια αναλαμβάνει το ρόλο του JobExecutor της άσκησης 2, δημιουργώντας πάντοτε 2 Workers με τα αντίστοιχα named pipes. Και εδώ, ο τερματισμός γίνεται με ομαλό τρόπο όπως και στο παραπάνω ερώτημα. Οι consumers, αναλαμβάνουν το ρόλο του «κατεβάσματος» της σελίδας από τον WebServer και στη συνέχεια τη σκανάρουν για άλλα links. Αν αυτά δεν υπάρχουν ήδη στο δίσκο και δε βρίσκονται στην ουρά, τότε τοποθετούνται στην ουρά. Κατά την απομάκρυνση στοιχείου από την ουρά, δημιουργείται τότε (ενώ ο mutex είναι ακόμα κλειδωμένος) ένα «άδειο» αρχείο, ώστε αν τύχει κατά το κατέβασμα και ένα άλλο νήμα βρει link για το ίδιο αρχείο, να μην το ξανατοποθετήσει στην ουρά. Έτσι, αφού το thread «κατεβάσει» το περιεχόμενο του (ως τώρα) άδειου αρχείου, το ανοίγει και το συμπληρώνει.

Η άσκηση ελέγχθηκε μέσω valgrind και δεν υπάρχει κανένα memory leak τόσο στον WebServer όσο και στον WebCrawler. Επίσης, εμπεριέχεται και Makefile που υλοποιεί separate compilation.

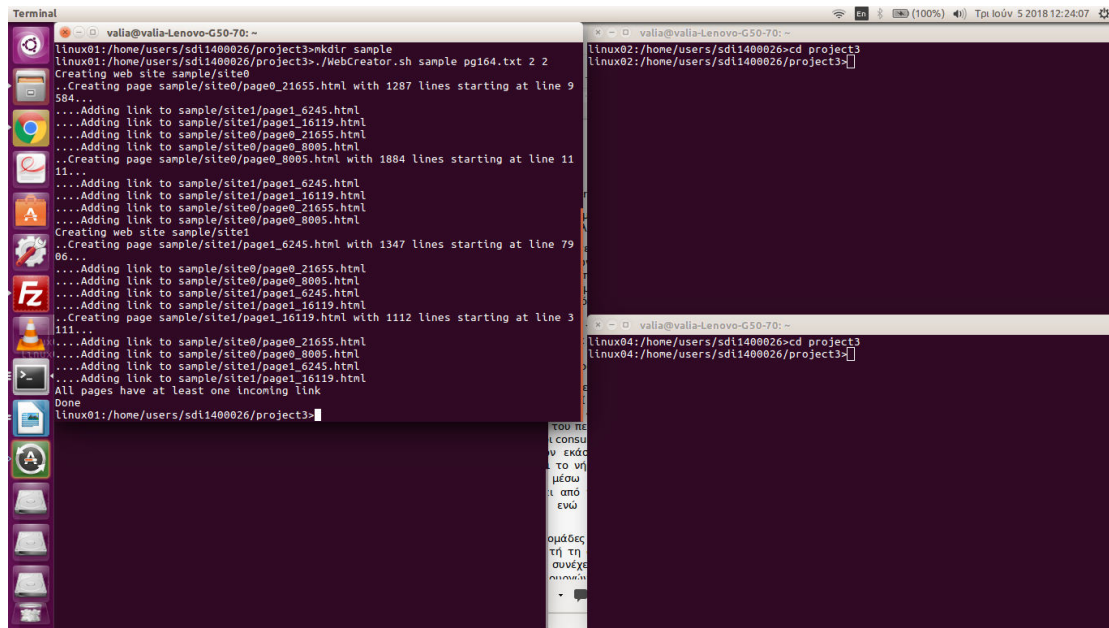
Τόσο στον WebServer όσο και στον WebCrawler, για κάθε κοινόχρηστη συνθήκη ελέγχου (π.χ. πλήρης buffer ή empty buffer) έχω βάλει αντίστοιχη μεταβλητή συνθήκης, ώστε να μην υπάρχει καθόλου busy waiting. Αντίστοιχα έχω βάλει και mutexes για την προστασία των κοινόχρηστων πόρων. Για τη υλοποίηση του WebServer και του WebCrawler, χρησιμοποίησα κώδικα από τις διαφάνειες του μαθήματος (κώδικας για server/client σε tcp sockets και κώδικας παραγωγού καταναλωτή με κυκλικό buffer).

Ενδεικτικές εκτελέσεις

Ανοίγω 4 terminal , και συνδέομαι σε διαφορετικό μηχάνημα.

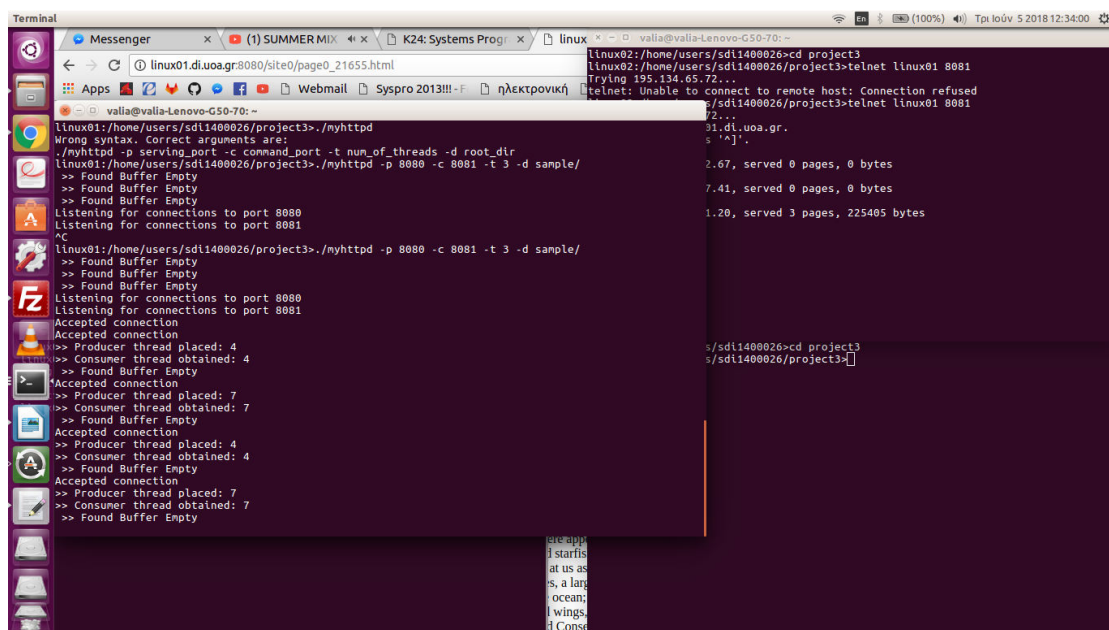
Σο πρώτο τρέχω το bash για τη δημιουργία των pages.

./WebCreator.sh sample pg164.txt 2 2 αφού κάνω mkdir sample.



```
linux01:/home/users/sdi1400026/project3>mkdir sample
linux01:/home/users/sdi1400026/project3>./WebCreator.sh sample pg164.txt 2 2
Creating web site sample/site0
..Creating page sample/site0/page0_21655.html with 1287 lines starting at line 9
584...
...Adding link to sample/site1/page1_6245.html
...Adding link to sample/site1/page1_16119.html
...Adding link to sample/site0/page0_21655.html
...Adding link to sample/site0/page0_8005.html
..Creating page sample/site0/page0_8005.html with 1884 lines starting at line 11
11...
...Adding link to sample/site1/page1_6245.html
...Adding link to sample/site1/page1_16119.html
...Adding link to sample/site0/page0_21655.html
...Adding link to sample/site0/page0_8005.html
..Creating web site sample/site1
..Creating page sample/site1/page1_6245.html with 1347 lines starting at line 79
06...
...Adding link to sample/site0/page0_21655.html
...Adding link to sample/site0/page0_8005.html
...Adding link to sample/site1/page1_6245.html
...Adding link to sample/site1/page1_16119.html
..Creating page sample/site1/page1_16119.html with 1112 lines starting at line 3
111...
...Adding link to sample/site0/page0_21655.html
...Adding link to sample/site0/page0_8005.html
...Adding link to sample/site1/page1_6245.html
...Adding link to sample/site1/page1_16119.html
All pages have at least one incoming link
Done
linux01:/home/users/sdi1400026/project3>
```

Υστερα, ./myhttpd -p 8080 -c 8081 -t 3 -d sample/



```
linux01:/home/users/sdi1400026/project3>./myhttpd
Wrong syntax. Correct arguments are:
./myhttpd -p serving_port -c command_port -t num_of_threads -d root_dir
linux01:/home/users/sdi1400026/project3>./myhttpd -p 8080 -c 8081 -t 3 -d sample/
>> Found Buffer Empty
>> Found Buffer Empty
Listening for connections to port 8080
Listening for connections to port 8081
^C
linux01:/home/users/sdi1400026/project3>./myhttpd -p 8080 -c 8081 -t 3 -d sample/
>> Found Buffer Empty
>> Found Buffer Empty
>> Found Buffer Empty
Listening for connections to port 8080
Listening for connections to port 8081
Accepted connection
Accepted connection
>> Producer thread placed: 4
>> Consumer thread obtained: 4
>> Found Buffer Empty
Accepted connection
>> Producer thread placed: 7
>> Consumer thread obtained: 7
>> Found Buffer Empty
Accepted connection
>> Producer thread placed: 4
>> Consumer thread obtained: 4
>> Found Buffer Empty
Accepted connection
>> Producer thread placed: 7
>> Consumer thread obtained: 7
>> Found Buffer Empty
```

Πηγαίνω σε έναν browser και πληκτρολογώ μια τυχαία σελίδα από αυτές που δημιουργήθηκαν:

[http://linux01.di.uoa.gr:8080/site0/page0\\_21655.html](http://linux01.di.uoa.gr:8080/site0/page0_21655.html)





```
Terminal
valla@valla-Lenovo-G50-70:~$
Welcome to Ubuntu 16.04.4 LTS (GNU/Linux 4.13.0-41-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

18 packages can be updated.
0 updates are security updates.

*** System restart required ***
Last login: Thu May 31 10:36:34 2018 from 2.86.88.80
linux03:/home/users/sdi1400026~$ clear
linux03:/home/users/sdi1400026~$ cd project3
linux03:/home/users/sdi1400026/project3~$ ./mycrawler
Wrong syntax. Correct arguments are:
./mycrawler -h host_or_IP -p port -c command_port -t num_of_threads -d save_dir startinlinux03:/home/users/sdi1400026/project3~$ ./mycrawler -h 8080 -c 8082 -t 3 -d saved /site0/page0_21655.html
Wrong syntax. Correct arguments are:
./mycrawler -h host_or_IP -p port -c command_port -t num_of_threads -d save_dir starting_URL
linux03:/home/users/sdi1400026/project3~$ ./mycrawler -h 8080 -c 8082 -t 3 -d saved /site0/page0_21655.html
Wrong syntax. Correct arguments are:
./mycrawler -h host_or_IP -p port -c command_port -t num_of_threads -d save_dir starting_URL
linux03:/home/users/sdi1400026/project3~$ clear

linux03:/home/users/sdi1400026/project3~$ ./mycrawler -h 8080 -c 8082 -t 3 -d saved /site0/page0_21655.html
Wrong syntax. Correct arguments are:
./mycrawler -h host_or_IP -p port -c command_port -t num_of_threads -d save_dir starting_URL
linux03:/home/users/sdi1400026/project3~$ ./mycrawler -h linux01 -p 8080 -c 8082 -t 3 -d saved /site0/page0_21655.html
>> Found Buffer Empty
>> Found Buffer Empty
Producer is waiting for files to be downloaded.
Consumer obtained: http://linux01:8080/site0/page0_21655.html
>> Found Buffer Empty
Connecting to linux01 port 8080
>> Found Buffer Empty
Producer is sending directories to Workers.
Listening for connections to port 8082
Producer is ready to receive telnet connections.
saved/site0
Worker with PID 21987 is waiting commands
Worker with PID 21986 is waiting commands
Accepted connection
```

Στο 4ο τερμινάλ:

```
Terminal
valla@valla-Lenovo-G50-70:~$
valla@valla-Lenovo-G50-70:~$ ssh -l sdi1400026 linux04.di.uoa.gr
sdi1400026@linux04.di.uoa.gr's password:
Welcome to Ubuntu 16.04.4 LTS (GNU/Linux 4.13.0-43-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

18 packages can be updated.
0 updates are security updates.

Last login: Tue Jun  5 12:06:14 2018 from 62.1.91.184
linux04:/home/users/sdi1400026~$ clear
linux04:/home/users/sdi1400026~$ cd project3
linux04:/home/users/sdi1400026/project3~$ telnet linux03 8082
Trying 195.134.65.74...
Connected to linux03.di.uoa.gr.
Escape character is '^['.
STATS
Crawler up for 01:27.62, downloaded 1 pages, 67844 bytes
SEARCH sea
linux03:/home/saved/site0/page0_21655.html 27 "The sea is open!!" was all he said.
aved /site0/saved/site0/page0_21655.html 53 and four yards. We made for it, but carefully, for the sea might be
Wrong syntax. Saved/site0/page0_21655.html 60 the sixteenth parallel, the sea is covered with floating ice of enormous
./mycrawler -s saved/site0/page0_21655.html 339 basalt. From thence we looked upon a vast sea which, towards the
ir starting unsaved/site0/page0_21655.html 396 ice increased on the open water. The sea seemed everywhere alike.
linux03:/home/saved/site0/page0_21655.html 405 cold, have emigrated to the north, these sea mammals remain sole
2 -t 3 -d saved/site0/page0_21655.html 871 water, drawn up from the sea after only having gone through the
>> Found Buffersaved/site0/page0_21655.html 1064 fuel and varech, of which the open polar sea contains so many
>> Found Buffersaved/site0/page0_21655.html 1075 recognised the following day. The depth of the sea was moderate. On
Producer is wsaved/site0/page0_21655.html 1076 the shores our nets brought in beautiful specimens of sea weed, and
Consumer obtasaved/site0/page0_21655.html 1226 the sea of the Antilles. April 16th, we sighted Martinique and
>> Found Buffersaved/site0/page0_21655.html 1233 the open sea it could not be thought of. The Canadian, Conseil, and I
Connecting to 2 out of 2 workers replied on time
>> Found Buff
Producer is se
Listening for
Producer is ready to receive telnet connections.
saved/site0
Worker with PID 21987 is waiting commands
Worker with PID 21986 is waiting commands
Accepted connection
```

και κάνω και μια search.

Και τέλος όταν γράψω SHUTDOWN

```
Terminal
* - - - valia@valia-Lenovo-G50-70: ~
>>> Found Buffer Empty
Accepted connection
>>> Producer thread placed: 7
>>> Consumer thread obtained: 7
>>> Found Buffer Empty
Accepted connection
>>> Producer thread placed: 4
>>> Consumer thread obtained: 4
>>> Found Buffer Empty
Accepted connection
>>> Producer thread placed: 7
>>> Consumer thread obtained: 7
>>> Found Buffer Empty
Accepted connection
>>> Producer thread placed: 4
>>> Consumer thread obtained: 4
>>> Found Buffer Empty
Consumer thread is exiting.
Consumer thread is exiting.
Consumer thread is exiting.
All threads exited normally, allocated memory is released.
linux01:/home/users/sdi1400026/project3>

* - - - valia@valia-Lenovo-G50-70: ~
linux03:/home/users/sdi1400026/project3>./mycrawler -h 8080 -c 8082 -t 3 -d s
aved /site0/page0_21655.html
Wrong syntax. Correct arguments are:
./mycrawler -h host_or_IP -p port -c command_port -t num_of_threads -d save_d
ir_starting_URL
linux03:/home/users/sdi1400026/project3>./mycrawler -h linux01 -p 8080 -c 808
2 -t 3 -d saved /site0/page0_21655.html
>>> Found Buffer Empty
>>> Found Buffer Empty
Producer is waiting for files to be downloaded.
Consumer obtained: http://linux01:8080/site0/page0_21655.html
>>> Found Buffer Empty
Connecting to linux01 port 8080
>>> Found Buffer Empty
Producer is sending directories to Workers.
Listening for connections to port 8082
Producer is ready to receive telnet connections.
saved/site0
Worker with PID 21987 is waiting commands
Worker with PID 21986 is waiting commands
Accepted connection
All threads exited normally, allocated memory is released.

* - - - valia@valia-Lenovo-G50-70: ~
*** System restart required ***
Last login: Tue Jun 5 12:05:58 2018 from 62.1.91.184
linux02:/home/users/sdi1400026>clear
linux02:/home/users/sdi1400026>cd project3
linux02:/home/users/sdi1400026/project3>telnet linux01 8081
Trying 195.134.65.72...
telnet: Unable to connect to remote host: Connection refused
linux02:/home/users/sdi1400026/project3>telnet linux01 8081
Trying 195.134.65.72...
Connected to linux01.dl.uoa.gr.
Escape character is '^['.
STATS
Server up for 00:12:67, served 0 pages, 0 bytes
STATS
Server up for 01:17:41, served 0 pages, 0 bytes
STATS
Server up for 03:11:20, served 3 pages, 225405 bytes
STATS
Server up for 21:15:58, served 4 pages, 293249 bytes
SHUTDOWN
Server is shutting down gently.
Connection closed by foreign host.
linux02:/home/users/sdi1400026/project3>

saved/site0/page0_21655.html 396 ice increased on the open water. The sea seemed e
verywhere alike.
saved/site0/page0_21655.html 405 cold, have emigrated to the north, these sea mamm
als remain sole
saved/site0/page0_21655.html 871 water, drawn up from the sea after only having go
ne through the
saved/site0/page0_21655.html 1064 fuci and varech, of which the open polar sea con
tains so many
saved/site0/page0_21655.html 1075 recognised the following day. The depth of the s
ea was moderate. On
saved/site0/page0_21655.html 1076 the shores our nets brought in beautiful specime
ns of sea weed, and
saved/site0/page0_21655.html 1226 the sea of the Antilles. April 16th, we sighted
Martinique and
saved/site0/page0_21655.html 1233 the open sea it could not be thought of. The Can
adian, Consell, and I
2 out of 2 workers replied on time
SHUTDOWN
Reply: 1
Reply: 0
All workers found 1 strings and exited.
Crawler is shutting down gently.
Connection closed by foreign host.
linux04:/home/users/sdi1400026/project3>
```