

Subject: Multi-thread Compressor – City University CS3103 Group Project

Group Number: 10

Members: Pao-Heng HSU (Code Execution), Yo-Che LEE (Design/Report), Haier LIN (Integration)

(The group feel like each member getting the same grade since we worked hard together)

The base of the program is the compressor we made in Assignment 1, with multi-thread concept applied on to it. Below are some features of our design:

1. To the program for efficient, one particularly way is to use memory-mapped files, available via `mmap()`. By mapping the input file into the address space, it can then access bytes of the input file via pointers making it quite efficient.
2. The parallel compression process can be treated as producer-consumer problem. The pzip uses one thread of producer to map files and multiple threads of consumers to compress pages. In this case, all the threads, including the main thread are treated as worker threads.
3. On load balancing, the main thread first maps the files then enter worker more. There are two main tasks that worker threads are in charge of, encoding and outputting. For each cycle, a worker thread will be designated of one task by the workloads.
 - a. Encoding: The maximum size of data a worker thread can encode is 5MB, which will be placed in an output array.
 - b. Outputting: Fetch the next output result from the output array then apply output.
4. The number of threads being created is “`get_nprocs()-1`”, which lead to a total of `get_nprocs()` with the main thread that equals to the core number of the computer it is running on, which is expected to enhance the efficiency as much as possible.

Applied concepts: `mmap`, `pthread`, `readdir`, `mutex(lock)`, condition variables.

Test Result

```
[ubt18a:/home/bsft15/yuenflam2/cs3103-project/project> make test
TEST 0 - clean build (program should compile without errors or warnings)
Test finished in 0.174 seconds
RESULT passed

TEST 1 - single file test, a small file of 10 MB (2 sec timeout)
Test finished in 0.021 seconds
RESULT passed

TEST 2 - multiple files test, twelve small files of 10 MB, 20 MB, 30 MB, 10 MB, 20 MB, 30 MB, 10 MB, 20 MB, 30 MB, 10 MB, 20 MB, 30 MB (2 sec timeout)
Test finished in 0.094 seconds
RESULT passed

TEST 3 - empty file test (2 sec timeout)
Test finished in 0.012 seconds
RESULT passed

TEST 4 - no file test (2 sec timeout)
Test finished in 0.005 seconds
RESULT passed

TEST 5 - single large file test, a large file of 100 MB (2 sec timeout)
Test finished in 0.025 seconds
RESULT passed

TEST 6 - multiple large files test, six large files of 100 MB, 200 MB, 300 MB, 100 MB, 200 MB, 300 MB (2 sec timeout)
Test finished in 0.183 seconds
RESULT passed

TEST 7 - directory test, a directory that contains twelve small files of 10 MB, 20 MB, 30 MB, 40 MB, 10 MB, 20 MB, 30 MB, 40 MB, 10 MB, 20 MB, 30 MB, 40 MB (2 sec timeout)
Test finished in 0.163 seconds
RESULT passed

TEST 8 - mixed test 1, a directory that contains six small files of 10 MB, 20 MB, 10 MB, 20 MB, 10 MB, 20 MB and six large files outside directory of 100 MB, 200 MB, 300 MB, 100 MB, 200 MB, 300 MB (2 sec timeout)
Test finished in 0.160 seconds
RESULT passed

TEST 9 - mixed test 2, a directory that contains six large files of 100 MB, 200 MB, 100 MB, 200 MB, 100 MB, 200 MB, and six small files outside directory of 30 MB, 40 MB, 30 MB, 40 MB, 30 MB, 40 MB (2 sec timeout)
Test finished in 0.156 seconds
RESULT passed

TEST 10 - mixed test 3, two directories that contain three small files of 10 MB, 20 MB, 10 MB and three large files of 200 MB, 100 MB, 200 MB, and six small files outside directory of 10 MB, 20 MB, 10 MB, 20 MB, 10 MB, 20 MB (2 sec timeout)
Test finished in 0.114 seconds
RESULT passed
```

```
[ubt18a:/home/bsft15/yuenflam2/cs3103-project/project> make run
TEST 1 - single file test, a small file of 10 MB (2 sec timeout)
Test finished in 0.019 seconds
RESULT passed

TEST 2 - multiple files test, twelve small files of 10 MB, 20 MB, 30 MB, 10 MB, 20 MB, 30 MB, 10 MB, 20 MB, 30 MB, 10 MB, 20 MB, 30 MB (2 sec timeout)
Test finished in 0.218 seconds
RESULT passed

TEST 3 - empty file test (2 sec timeout)
Test finished in 0.012 seconds
RESULT passed

TEST 4 - no file test (2 sec timeout)
Test finished in 0.005 seconds
RESULT passed

TEST 5 - single large file test, a large file of 100 MB (2 sec timeout)
Test finished in 0.025 seconds
RESULT passed

TEST 6 - multiple large files test, six large files of 100 MB, 200 MB, 300 MB, 100 MB, 200 MB, 300 MB (2 sec timeout)
Test finished in 0.158 seconds
RESULT passed

TEST 7 - directory test, a directory that contains twelve small files of 10 MB, 20 MB, 30 MB, 40 MB, 10 MB, 20 MB, 30 MB, 40 MB, 10 MB, 20 MB, 30 MB, 40 MB (2 sec timeout)
Test finished in 0.103 seconds
RESULT passed

TEST 8 - mixed test 1, a directory that contains six small files of 10 MB, 20 MB, 10 MB, 20 MB, 10 MB, 20 MB and six large files outside directory of 100 MB, 200 MB, 300 MB, 100 MB, 200 MB, 300 MB (2 sec timeout)
Test finished in 0.165 seconds
RESULT passed

TEST 9 - mixed test 2, a directory that contains six large files of 100 MB, 200 MB, 100 MB, 200 MB, 100 MB, 200 MB, and six small files outside directory of 30 MB, 40 MB, 30 MB, 40 MB, 30 MB, 40 MB (2 sec timeout)
Test finished in 0.144 seconds
RESULT passed

TEST 10 - mixed test 3, two directories that contain three small files of 10 MB, 20 MB, 10 MB and three large files of 200 MB, 100 MB, 200 MB, and six small files outside directory of 10 MB, 20 MB, 10 MB, 20 MB, 10 MB, 20 MB (2 sec timeout)
Test finished in 0.135 seconds
RESULT passed
```