Contents

1 README

```
ransha
Ran Shaham (203781000)
EX: 3

FILES:
README            -- This file
MapReduceFramework.cpp    -- My implementation of the uthreads library.
Search.cpp        -- The test (Task 1) for the library.
Makefile        -- Creates the library and 'Search' when called with
                no arguments
my_pthread.h        -- Wrapper for some of the pthread functions to handle
                unsuccessful pthread calls
error_handle.h        -- A header that includes the function that handles
                errors in this exercise
RR.jpg            -- Gantt chart for question 6 - Round Robin
FCFS.jpg        -- " - First Come First Serve
SRTF.jpg        -- " - Shortest Remaining Time First
Priority.jpg        -- " - Priority Scheduling


REMARKS:
Framework Design:
Most of the design was driven directly from the exercise description, except
for the following:
* The ExecMap output data structures are vectors, which allow the shuffle to
  read from them while the map writes (push_back) to them.
* Synchronization between the map and the shuffle is done by signaling the
  shuffle thread when a new chunk is written.
* The shuffle keeps a vector of indices. Each entry specifies the last index
  in each ExecMap vector that has been written, and DOES NOT pop items from
  the map's vector. After reading, it updates the index for that container.

Search.cpp Design:
First I'll describe the k1-k3,v1-v3 types I chose-
* The <k1,v1> pairs are the directory names and the substring to search
  respectively. This way, on each call to the Map function, the given directory
  is opened (it is written in the forums that we can assume that each folder
  will appear only once in the arguments), iterated over and searched for file
  names that contain the substring.
* The k2 is the filename that contains the substring. That is, a class that
  has a string member that holds the filename. The v2's only purpose is to
  free allocated k2 pointers. It takes a k2* as an argument and when destructed
  it also deletes the k2 that it holds. The idea is that every v2 created has
  its own k2, and the shuffle can "ignore" (merge) several <k2,v2> pairs to:
  <k2, list<v2>> pairs - i.e. no v2 is ever ignored. Therefore when the v2s
  are destroyed, all their k2s can be safely destroyed, and that happens
  implicitly when calling to "delete v2" (destructor). This way all of the
  <k2,v2> allocated memory is freed.
* k3 is the same as k2 (filename), and v3 is simply the number of times the
  file name that is held in k3 appeared in the input directories. For example,
  if the directory structre is this:

  root    -> dir1    -> a, b, c, .vimrc
    -> dir2    -> d, a, e
    -> dir3    -> Makefile, README, test.o, a
    -> dir4    -> .ssh, .config, Games

  then calling "Search a dir1 dir2 dir3" will yield [after reduce] the <k3,v3>
  entry: <"a",3> (among the other entries - <"e",1>, <"b",1> etc.).
```

Next, the Map/Reduce operations:
* Map - Gets a <dirname, strToFind> pair. It checks whether the 'dirname' is a valid directory (and if it exists) - if not it returns. Otherwise, it opens the directory and for each entry (file/link/dir) checks if the 'strToFind' is a substring of that entry's name. If so, it creates the <filename, v2> pair using (using the 'new' operator) and sends it to Emit2.
* Reduce - Gets a <filename, list<v2>> pair. It copies the filename to the k3 entry, and sets the value of v3 (filename count) to the size of the list of v2s. This works because the shuffle merges all files with the same name while keeping the v2s in a list for that filename, so the lise size is exactly the number of occurunces of the filename.
  Then, it iterates over the input list and delete the v2* values (and k2, implicitly), since they are no longer needed for the framework.

Having said that, the Search procedure is as follows:

1. Prepare the input - convert the input string arguments to the string to search and the directory names, and create a list of <k1*, v1*> pairs.
2. Run the framework and get the output <filename,count> list.
3. For each list entry of <file_i, count_i>, print file_i count_i times.
4. Iterate over the input and output lists and delete allocated memory

ANSWERS:

Q1:
The program will work with multiple processes (instead of threads).
Each process
* The process using select is the Shuffle, since it waits for data to be written

Q2:
Since the user's computer can run 8 threads concurrently - 1 on each core, multiThreadLevel should be at least 7 which will yield 8 threads running at the same time: 7 ExecMap threads + 1 shuffle thread.
Since the computer doesn't support hyper-threading, any additional thread will wait till a thread that was created earlier will finish running, therefore it is redundant. E.g., say multiThreadLevel is 10, then 8 ExecMap threads are created and being run and the remaining 2 WAIT until 2 of the earlier ones finish running. When they do, by definition there is no more mapping to be done so they are immidiately terminated. Then, after the mapping is finished can a shuffle be created and only start shuffling.
On the other hand, less than 7 won't utilze all system resources - after creating 4 (for example) ExecMap threads the main thread then creates one shuffle thread and waits for them to finish. So there will be 5 working thread untill the shuffle is done and 2 cores will remain unused this entire time (lower CPU utilization).
In conclusion, I'd suggest to set multiThreadLevel to 7.

Q3:
a. Nira's solution - Single thread
* Does not utilze multi-cores since only one thread of execution is running.
* No scheduler is needed since no concurrency is happening, the program flow is serial.
* No communication is needed since there's a single thread and process.
* When waiting for I/O operations or other time consuming actions, the program does not progress
* The overall speed is probably slower than the other methods, since the process can spend a lot of time waiting (for disk access when writing the log, for example) and cannot do anything in this time.

b. Moti - POSIX's library
* Utilizes multi-cores since the OS is aware of kernel-level threads and can run a thread for each core or processor concurrently.
* The library manages its own scheduling methods so besides block / signal there is no much scheduling that can be done.
* Communication time is relatively fast since the threads share heap and static data segments, and can read from them concurrently. Writing requires locking shared resources, but it is still better than multi-process communication.

* When one thread is blocked, another can proceed, although when accessing disk
  no concurrency can occur (that is, two threads cannot acess the disk in the
  same time).
* The overall speed depends on whether the machine running the framework has
  multiple cores or not. If not, there is no advantage for this method over
  user-level threads, and the overhead of context switching, thread creation
  and termination is bigger than in user-level threads.
  If the machine has multiple cores or processes, this method allows more than
  one thread running in the same time, thus is probably faster than the other
  methods.

c. Danny - User-level threads
* Since the OS is not aware of the multiple threads running, it does not
  run threads on multiple cores concurrently, so this method does not utilize
  multiple cores.
* This is THE method for creating scheduler based on internal data - the user
  can do whatever he/she wants, and has access to all data in the program, so
  this method is very flexible in terms of scheduling.
* Communication time is similar to kernel-level threads. If no built-in OS
  mutex types are used (which gurantee atomic operations) synchronization can
  be more tricky than in kernel-level.
* The OS is not aware of the fact multiple threads are running, so when the
  process makes a system call, the OS is likely to block it, therefore not
  allowing other user-level threads to run in the meantime. If the OS does
  not block the entire process, then the program can progress on another
  thread while another one waits.
* This is likely to be a fast implementation on single core machines, since
  context switches and thread actions require mainly function calls, and not
  system calls, therefore overhead time is reduced. On multiple cores, no
  concurrency is involved so this method is likely to be slower than methods
  that enable concurrency (multi-process or kernel-level threads).

d. Galit - Multi-processes
* This method utilizes multi-cores since modern OSs tend to do so with
  multiple processes.
* The scheduling is almost entirely up to the OS to decide - there are some
  hints that can affect the scheduler (nice values, for example) but are based
  on general process role, and not fine-grained internal data.
* Communication is trickier than in other methods - the processes need pipes
  or files to communicate and handling files require sys-calls, which are slow
  as we measured in EX1.
* Each process is more or less independent, so while one is blocked another can
  proceed (if it should, logically).
* In my opinion (i.e - based on my knowledge so far, not experience) this
  method will be slower than the others on single core machines, and perhaps
  a bit faster on multi-core than the single-thread or user-level threads
  methods, but slower than the kernel-level threads.

Q4:
Processes:
a. Stack is not shared, each process gets its own stack.
b. Same for the heap.
c. Same for global variables.
When using fork() sys call, the above data segments are copied to the child
process, not shared.

Kernel-level threads:
a. Each thread gets its own stack
b. Threads share heap data, since they reside in the same process
c. Same as heap, the process' static data segment is the same for all threads

User-level threads:
Exactly the same as ernel-level threads.
A stack can't be shared between threads since they run different functions and
execute a different set of commands (perhaps), so each thread should have its
own stack for local variables and adresses.
The heap and static data segments are shared, since the threads are run in a
one process.

Q5:
A deadlock is a situation in which two or more concurrent running threads of execution (processes/ threads) try to access some shared resources; each of them blocks another and is blocked by another, thus neither of them progresses with their execution.
For example, threads A,B both need resources x,y to perform an operation. Consider the following scenerio: A runs, reserves x; then, a context switch occurs and B reserves y. If it tries to reserve x it will be blocked. If A tries to reserve y it will also be blocked. Therefore they both block each other and neither progresses.

A livelock is also a situation in which two or more running threads don't proceed with their execution. In this case, all threads are taking active actions in order to let the other threads finish their work, and then proceed. If each thread lets the other threads proceed and no thread actually proceeds, a livelock occurs.
For example, say threads A,B need resources x,y as before. Again, A reserves x and B reserves y. Now B tries to reserve x and fails so it lets go of y and waits 1 second. Before it does, A gets the control and tries to access y, which is still in B's control, so it lets go of x and waits for 1 second. Then, they try again using the same method. This creates the above scenerio, where 2 threads aren't progressing but both are taking active actions to prevent locks.

Q6:

# NAME	Avg. wait time (format: (P1:a+b+... + P2:c+d+... + ...) / 5 = result
1. RR	(1+2+3+2 + 1 + 2 + 2+3+2 + 3) / 5 = 4.2
2. FCFS	(0 + (10-1) + (11-3) + (13-7) + (25-8)) / 5 = 8
3. SRTF	(1+2+1 + 0 + 0 + 7 + 0) / 5 = 2.2
4. PRIO	(0 + (10-1) + (23-3) + (11-7) + (25-8)) / 5 = 10

# NAME	Turnaround time
1. RR	(18 + (3-1) + (7-3) + (26-7) + (12-8)) / 5 = 9.4
2. FCFS	(10 + (11-1) + (13-3) + (25-7) + (26-8)) / 5 = 13.2
3. SRTF	(14 + (2-1) + (5-3) + (26-7) + (9-8)) / 5 = 7.4
4. PRIO	(10 + (11-1) + (25-3) + (23-7) + (26-8)) / 5 = 15.2

* (All time calculations are ignoring context switching)

2 Appeal

```
My Errors:

* First, I accidently submitted an old tar in which one major problem was yet
  to be fixed. That is, I already made the change (prior to submission), and
  submitted the unchanged file ( :-( )
  The problem was segfault due to uninitialized ExecReduce data structures,
  that was fixed using a simple mutex barrier to prevent the ExecReduce
  threads from running until their data structures are initialized - exactly
  the same as the ExecMap problem that was discussed in the forums.
  This is a stupid-reckless mistake that failed probably 90% of the tests.

* My second error was a tricky one. I tried to avoid a situation in which the
  shuffle blocks the ExecMaps frequently, while reading their output.
  Therefore I used a vector of ExecMap results (for each thread) to which the
  ExecMap pushed new data. The shuffle read it from the beggining while
  keeping track of the index it read (without locking it, since we're only
  reading and not modifying). The problem is when the shuffle tried to read
  from this vector in the same time the ExecMap pushed data to it, and this
  data adding invoked an internal vector's data reallocation, it attempted to
  access invalidated data - and that resulted in a segfault.
  Note that a simple 'realloc' wouldn't have invalidate the vector's data -
  it happened only when a full blown 'free' + 'malloc' was performed.
  This happened only when the maps pushed a huge amount of data -
  I tested my library with big files and complex situations - and didn't
  encounter this problem. Only running the WordCount test did on very big files
  did.

  This was fixed using an array (vector) of mutexes for each ExecMap, which is
  locked whenever the ExecMap pushes data to the vector (and unlocked after
  that), and when the shuffle tries to access this ExecMap's data.
```

3 FCFS.jpg

4 Makefile

```
1  CFLAGS=-std=c++11 -Wall -Wextra -g -pthread $(INCS)
2
3  # cpp to object files rule
4  %.o: %.cpp
5      $(CXX) $(CFLAGS) -c $<
6
7  # Library Stuff
8  INCS=-I.
9  LOADLIBES=-L.
10
11 LIBSRC=MapReduceFramework.cpp
12 LIBH=MapReduceFramework.h
13 LIBOBJ=MapReduceFramework.o
14 ARFLAGS=rvs
15 RANLIB=ranlib
16
17 LIBTARGET=MapReduceFramework.a
18
19 # test rules
20 TEST_SRC=Search.cpp
21 TEST_FILE=Search
22 VALGRIND_FLAGS = --leak-check=full --show-possibly-lost=yes \
23          --show-reachable=yes --undef-value-errors=yes
24
25 $(TEST_FILE): $(TEST_SRC) $(LIBTARGET)
26     $(CXX) $(CFLAGS) $(LOADLIBES) $^ -o $@
27
28 ValgrindTest: $(TEST_OBJ)
29     valgrind $(VALGRIND_FLAGS) ./$<
30
31
32 # library rules
33 $(LIBTARGET): $(LIBOBJ)
34     $(AR) $(ARFLAGS) $@ $^
35     $(RANLIB) $@
36
37 $(LIBOBJ): $(LIBSRC)
38     $(CXX) $(CFLAGS) -c $< -o $(LIBOBJ)
39
40 # cleaning and such
41 RM=rm -fv
42 LOG_FILE=.MapReduceFramework.log
43 clean:
44     $(RM) $(TEST_FILE) $(LIBTARGET) *.o $(LOG_FILE)
45
46 depend:
47     makedepend -- $(CFLAGS) -- $(SRC) $(LIBSRC)
48
49 # tar rule
50 TAR=tar
51 TARFLAGS=-cvf
52 TARNAME=ex3.tar
53 EXTRA_HEADERS=my_pthread.h error_handle.h
54 GANTT_PICS=RR.jpg FCFS.jpg SRTF.jpg Priority.jpg
55 TARSRCS=$(LIBSRC) $(TEST_SRC) Makefile README $(EXTRA_HEADERS) $(GANTT_PICS)
56
57 tar:
58     $(TAR) $(TARFLAGS) $(TARNAME) $(TARSRCS)
59
```

```
60  all: $(TEST_FILE)
61
62  .PHONY: all clean tar ValgrindTest
```

5 MapReduceFramework.cpp

```cpp
/* == Includes == */
#include "MapReduceFramework.h"

#include "my_pthread.h"        /* inlcudes pthread.h + error handling */
#include <fstream>
#include <vector>
#include <queue>
#include <map>

#include <cmath>          /* for fmin */
#include <sys/time.h>         /* gettimeofday */
#include <ctime>         /* for strftime */
#include <chrono>         /* time measurements */
#include <algorithm>         /* for std::move */

/* To reduce line-length and increase readability */
using std::vector;
using std::queue;
using std::map;
using std::multimap;
using std::list;
using std::pair;

/* == Some constants == */
#define CHUNK_SIZE 10
#define USEC_TO_NSEC(x) ((x) * 1000)
#define NANO_IN_SEC 1000000000
#define TIME_TO_WAIT 1000000

#define TIME_FORMAT "[%d.%m.%Y %T]"

#define LOG_FILENAME ".MapReduceFramework.log"
#define LOG_OPEN_MODE std::ofstream::out | std::ofstream::app

#define THREAD_EXECMAP "ExecMap"
#define THREAD_SHUFFLE "Shuffle"
#define THREAD_EXECREDUCE "ExecReduce"

#define MSG_FRAMEWORK_START(t) "runMapReduceFramework started with " << t \
    << " threads"
#define MSG_FRAMEWORK_END "runMapReduceFramework finished"
#define MSG_THREAD_CREATED(th) "Thread " th " created "
#define MSG_THREAD_TERMINATED(th) "Thread " th " terminated "
#define MSG_MAP_SHUFFLE_TIME(t) "Map and Shuffle took " << t << " ns"
#define MSG_REDUCE_TIME(t) "Reduce took " << t << " ns"

/* Struct for the map procedure. nextToRead is the only variable that *
 * will be mutexed, as specified in the instructions. */
typedef struct MapData
{
    MapReduceBase &mapReduce;
    vector<IN_ITEM> &itemsList;
    size_t numOfItems;
    size_t nextToRead;
    std::ofstream &logofs;
} MapData;

/* Same for reduce, the items list is static and therefore not in here. */
typedef struct ReduceData
```

```
{
    MapReduceBase &mapReduce;
    size_t threadNum;
    size_t numOfItems;
    size_t nextToRead;
    std::ofstream &logofs;
} ReduceData;

/* Comparator to sort keys by values, instead of adresses */
template<class T> struct ptr_less
{
    bool operator()(T* lhs, T* rhs)
    {
    return *lhs < *rhs;
    }
};

/* Threads variables */
vector<pthread_t> threadList; /* A list of existing ExecMap/Reduce threads */
pthread_t shuffleThread;      /* The shuffle thread */

/* ExecMap variables */
typedef vector<pair<k2Base*, v2Base*>> K2_V2_LIST;
/* a list of pairs in which the first element is the pthread_t (id) and *
 * the second is that thread's list (pointer) of the map-function-outputs */
vector<pair<pthread_t, K2_V2_LIST*>> mapResultLists;

/* Shuffle Variables */
/* An array of indices that specify till what index in the map-lists *
 * the shuffle thread has already read */
vector<size_t> shuffleCurrIndex;
/* Note that the map compares keys using ptr_less defined above */
map<k2Base*, V2_LIST*, ptr_less<k2Base>> shuffleOut;

/* Reduce Variables */
vector<pair<k2Base*, V2_LIST*>> reduceIn;
typedef queue<pair<k3Base*, v3Base*>> K3_V3_LIST;
/* Same as before, but for the reduce-function-outputs */
vector    <pair<pthread_t, K3_V3_LIST*>> reduceResultLists;
multimap<k3Base*, v3Base*, ptr_less<k3Base>> finalOutputMap;

/* Mutexes */
vector<pthread_mutex_t> mutex_map_write; // <--
pthread_mutex_t mutex_er_list;        /* barrier to init reduce dasts */
pthread_mutex_t mutex_map_read;        /* protect input list */
pthread_mutex_t mutex_em_map;        /* barrier to init maps */
pthread_mutex_t mutex_more_to_shuffle;    /* condition of shuffle loop */
pthread_mutex_t mutex_reduce_read;    /* protect reduce input */
pthread_mutex_t mutex_log_write;    /* protect writing to log */
/* Condition Variables */
pthread_cond_t cv_more_to_shuffle;
bool more_to_shuffle;


/**
 * Gets the current time as a string in the exercise format. i.e
 * "[DD.MM.YYYY HH:MM:SS]"
 */
string getTimeStr()
{
    size_t MAX_CHARS = 64;
    time_t rawtime;
    struct tm * timeinfo;

    time(&rawtime);
    timeinfo = localtime(&rawtime);
    char buff[MAX_CHARS];

```

```
128     strftime(buff, MAX_CHARS, TIME_FORMAT, timeinfo);
129     return buff;
130 }
131
132
133 /**
134  * Initializes all mutexes, cv and relevant static variables for the library.
135  */
136 void init(int multiThreadLevel)
137 {
138     // Pthread types
139     mutex_map_write.resize(multiThreadLevel); // <--
140     for (int i = 0; i < multiThreadLevel; ++i)
141         my_pthread_mutex_init(&mutex_map_write[i], nullptr);
142     my_pthread_mutex_init(&mutex_er_list, nullptr);
143     my_pthread_mutex_init(&mutex_map_read, nullptr);
144     my_pthread_mutex_init(&mutex_em_map, nullptr);
145     my_pthread_mutex_init(&mutex_more_to_shuffle, nullptr);
146     my_pthread_cond_init(&cv_more_to_shuffle, nullptr);
147     my_pthread_mutex_init(&mutex_reduce_read, nullptr);
148     my_pthread_mutex_init(&mutex_log_write, nullptr);
149     // Variables and data structures
150     more_to_shuffle = true;
151     threadList.resize(multiThreadLevel);
152     mapResultLists.resize(multiThreadLevel);
153     shuffleCurrIndex.resize(multiThreadLevel);
154     reduceResultLists.resize(multiThreadLevel);
155 }
156
157 /**
158  * Destroys mutexes, cv and clear the static data structures in use.
159  */
160 void cleanup()
161 {
162     // Pthread types
163     for (size_t i = 0; i < mutex_map_write.size(); ++i) // <--
164         my_pthread_mutex_destroy(&mutex_map_write[i]);
165     mutex_map_write.clear();
166     my_pthread_mutex_destroy(&mutex_er_list);
167     my_pthread_mutex_destroy(&mutex_map_read);
168     my_pthread_mutex_destroy(&mutex_em_map);
169     my_pthread_mutex_destroy(&mutex_more_to_shuffle);
170     my_pthread_cond_destroy(&cv_more_to_shuffle);
171     my_pthread_mutex_destroy(&mutex_reduce_read);
172     my_pthread_mutex_destroy(&mutex_log_write);
173     // heap-allocated data delete
174     for (size_t i = 0; i < mapResultLists.size(); ++i)
175     {
176         delete mapResultLists[i].second;
177         delete reduceResultLists[i].second;
178         mapResultLists[i].second = nullptr;
179         reduceResultLists[i].second = nullptr;
180     }
181     for (auto it = reduceIn.begin(); it != reduceIn.end(); ++it)
182     {
183         delete it->second;
184         it->second = nullptr;
185     }
186     // Variable & data structures reset
187     shuffleThread = 0;
188     threadList.clear();
189     mapResultLists.clear();
190     reduceResultLists.clear();
191     shuffleCurrIndex.clear();
192     shuffleOut.clear();
193     reduceIn.clear();
194     finalOutputMap.clear();
195 }
```

```
196
197 /**
198  * The ExecMap procedure: Reads data from the input <k1,v1> list and uses
199  * the user provided Map function to create a list of <k2,v2> values (for
200  * each thread running this function).
201  * Takes a MapData struct pointer as an argument.
202  */
203 void* ExecMap(void *arg)
204 {
205     /* A barrier, opens when all ExecMap data structures are *
206      * initialized */
207     my_pthread_mutex_lock(&mutex_em_map);
208     my_pthread_mutex_unlock(&mutex_em_map);
209     MapData* mapdata = (MapData*) arg;
210     size_t myItems; // The start index for this thread's data
211     // While there's more data to read, read a chunk and map.
212     while ((myItems = mapdata->nextToRead) < mapdata->numOfItems)
213     {
214         // Lock the index counter
215         my_pthread_mutex_lock(&mutex_map_read);
216         // Reserve items to map and get the start index of them.
217         myItems = (mapdata->nextToRead += CHUNK_SIZE) - CHUNK_SIZE;
218         // Unlock
219         my_pthread_mutex_unlock(&mutex_map_read);
220         // Map this thread's data
221         int end = fmin(myItems + CHUNK_SIZE, mapdata->numOfItems);
222         for (int i=myItems; i < end; ++i)
223         {
224             mapdata->mapReduce.Map(mapdata->itemsList.at(i).first,
225                         mapdata->itemsList.at(i).second);
226         }
227         // Notify the shuffle thread
228         my_pthread_cond_signal(&cv_more_to_shuffle);
229     }
230     my_pthread_mutex_lock(&mutex_log_write);
231     mapdata->logofs << MSG_THREAD_TERMINATED(THREAD_EXECMAP)
232         << getTimeStr() << endl;
233     my_pthread_mutex_unlock(&mutex_log_write);
234     return nullptr;
235 }
236
237
238 /**
239  * The Shuffle procedure: Merge <k2,v2> pairs from the ExecMap containers
240  * to <k2, queue<v2>> containers.
241  */
242 void* Shuffle(void *)
243 {
244     struct timespec timeToWake;
245     struct timeval timeNow;
246     pair<k2Base*, v2Base*> currPair;
247     k2Base* currKey;
248     v2Base* currVal;
249     auto foundList = shuffleOut.end();
250     V2_LIST* newList = nullptr;
251     K2_V2_LIST* currList = nullptr;
252     size_t currIdx = 0;
253
254     my_pthread_mutex_lock(&mutex_more_to_shuffle);
255     while (more_to_shuffle)
256     {
257         if (gettimeofday(&timeNow, nullptr) < 0)
258         {
259             handleError("gettimeofday");
260         }
261         timeToWake.tv_sec = timeNow.tv_sec +
262             (USEC_TO_NSEC(timeNow.tv_usec) + TIME_TO_WAIT) /
263                         NANO_IN_SEC;
```

```
264             timeToWake.tv_nsec = fmod(USEC_TO_NSEC(timeNow.tv_usec) +
265                 TIME_TO_WAIT, NANO_IN_SEC);
266             // Wait for data (or time passing)
267             my_pthread_cond_timedwait(&cv_more_to_shuffle,
268                     &mutex_more_to_shuffle, &timeToWake);
269             // Check every ExecMap container for items to shuffle
270             for (size_t i = 0; i < mapResultLists.size(); ++i)
271             {
272                 my_pthread_mutex_lock(&mutex_map_write[i]);
273                 currList = mapResultLists[i].second; // After locking
274                 // Read all unread data and shuffle it..
275                 while (!currList->empty())
276                 { // Now I actually modify the map results
277                     currPair = currList->back();
278                     currList->pop_back();
279                     currKey = currPair.first;
280                     currVal = currPair.second;
281                     foundList = shuffleOut.find(currKey);
282                     // If no k2 entry exists, create it.
283                     if (foundList == shuffleOut.end())
284                     {
285                         newList = new(std::nothrow) V2_LIST;
286                         // If new operator has failed
287                         if (newList == nullptr)
288                         {
289                             handleError("new operator");
290                         }
291                         newList->push_back(currVal);
292                         shuffleOut.insert(std::make_pair(
293                                     currKey,
294                                     newList));
295                         newList = nullptr;
296                     }
297                     else
298                     {
299                         foundList->second->push_back(currVal);
300                     }
301                 }
302                 // 'Mark' all read data as read
303                 shuffleCurrIndex[i] = currIdx;
304                 my_pthread_mutex_unlock(&mutex_map_write[i]);
305             }
306         }
307         my_pthread_mutex_unlock(&mutex_more_to_shuffle);
308         return nullptr;
309     }
310
311
312     /**
313      * The Reduce procedure: Read input from a data structure and run the reduce
314      * function on it to produce the output data
315      */
316     void* ExecReduce(void *arg)
317     {
318         my_pthread_mutex_lock(&mutex_er_list);
319         my_pthread_mutex_unlock(&mutex_er_list);
320         ReduceData* reducedata = (ReduceData*) arg;
321         size_t myItems; // The start index for this thread's data
322         // While there's more data to read, read a chunk
323         while ((myItems = reducedata->nextToRead) < reducedata->numOfItems)
324         {
325             // Lock the index counter
326             my_pthread_mutex_lock(&mutex_reduce_read);
327             // Reserve items to map and get the start index of them.
328             myItems = (reducedata->nextToRead += CHUNK_SIZE) - CHUNK_SIZE;
329             // Unlock
330             my_pthread_mutex_unlock(&mutex_reduce_read);
331             int end = fmin(myItems + CHUNK_SIZE, reducedata->numOfItems);
```

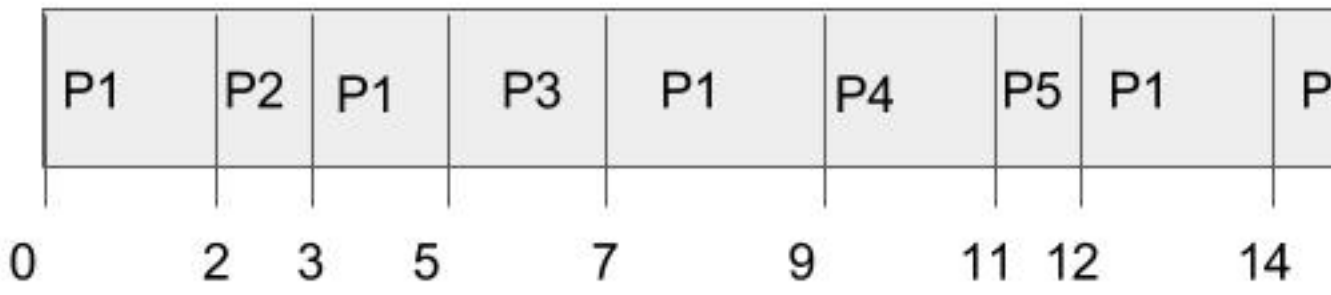
```
332         for (int i = myItems; i < end; ++i)
333         {
334             reducedata->mapReduce.Reduce( reduceIn.at(i).first,
335                             *reduceIn.at(i).second);
336         }
337     }
338     my_pthread_mutex_lock(&mutex_log_write);
339     reducedata->logofs << MSG_THREAD_TERMINATED(THREAD_EXECREDUCE)
340         << getTimeStr() << endl;
341     my_pthread_mutex_unlock(&mutex_log_write);
342     return nullptr;
343 }
344
345
346 /**
347  * Transforms the many reduce-thread-data-structures to one big map.
348  * Same as shuffle, but for reduce.
349  */
350 void finalize_output()
351 {
352     pair<k3Base*, v3Base*> currPair;
353     for (size_t i = 0; i < reduceResultLists.size(); ++i)
354     {
355         while( reduceResultLists[i].second != nullptr &&
356               !reduceResultLists[i].second->empty() )
357         {
358             currPair = reduceResultLists[i].second->front();
359             reduceResultLists[i].second->pop();
360             finalOutputMap.insert(currPair);
361         }
362     }
363 }
364
365
366 /**
367  * The main framework function.
368  */
369 OUT_ITEMS_LIST runMapReduceFramework(MapReduceBase &mapReduce,
370         IN_ITEMS_LIST &itemsList, int multiThreadLevel)
371 {
372     // open the log file, write and start measuring time
373     std::ofstream ofs(LOG_FILENAME, LOG_OPEN_MODE);
374     ofs << MSG_FRAMEWORK_START(multiThreadLevel) << endl;
375     std::chrono::high_resolution_clock::time_point t1 =
376         std::chrono::high_resolution_clock::now();
377
378     // Initialize mutexes, cvs, datastructures...
379     init(multiThreadLevel);
380     // Turn the in_items_list into a vector for random access
381     vector<IN_ITEM> itemsVector( std::begin(itemsList),
382                      std::end(itemsList)  );
383
384     MapData mapdata = {.mapReduce = mapReduce, .itemsList = itemsVector,
385         .numOfItems = itemsVector.size(),
386         .nextToRead = 0, .logofs = ofs};
387
388     // Create the ExecMap threads!
389     my_pthread_mutex_lock(&mutex_em_map); // Barrier to prevent sigsegv
390     for (int i = 0; i < multiThreadLevel; ++i)
391     {
392         my_pthread_mutex_lock(&mutex_log_write);
393         ofs << MSG_THREAD_CREATED(THREAD_EXECMAP) << getTimeStr()
394             << endl;
395         my_pthread_mutex_unlock(&mutex_log_write);
396         my_pthread_create(&threadList[i], nullptr, &ExecMap, &mapdata);
397     }
398     // Initialize all threads' maps
399     for (int i = 0; i < multiThreadLevel; ++i)
```

```
400         {
401             mapResultLists[i] = std::make_pair(threadList[i],
402                     new(std::nothrow) K2_V2_LIST);
403             if (mapResultLists[i].second == nullptr)
404             {
405                 handleError("new operator");
406             }
407         }
408         my_pthread_mutex_unlock(&mutex_em_map); // Remove the barrier
409 
410         // Create the shuffle thread
411         my_pthread_mutex_lock(&mutex_log_write);
412         ofs << MSG_THREAD_CREATED(THREAD_SHUFFLE) << getTimeStr() << endl;
413         my_pthread_mutex_unlock(&mutex_log_write);
414         my_pthread_create(&shuffleThread, nullptr, &Shuffle, nullptr);
415 
416         // Wait for the ExecMaps to finish
417         for (int i = 0; i < multiThreadLevel; ++i)
418         {
419             my_pthread_join(threadList[i], nullptr);
420         }
421         // Wait for the shuffle to finish
422         my_pthread_mutex_lock(&mutex_more_to_shuffle);
423         more_to_shuffle = false;
424         my_pthread_mutex_unlock(&mutex_more_to_shuffle);
425         my_pthread_cond_signal(&cv_more_to_shuffle);
426         my_pthread_join(shuffleThread, nullptr);
427 
428         // Stop time measuring
429         std::chrono::high_resolution_clock::time_point t2 =
430             std::chrono::high_resolution_clock::now();
431         auto mapShuffleDuration = std::chrono::duration_cast
432             <std::chrono::nanoseconds>(t2 - t1).count();
433         ofs << MSG_THREAD_TERMINATED(THREAD_SHUFFLE) << getTimeStr() << endl;
434 
435         // Transform the shuffle output map to a vector.
436         reduceIn.resize(shuffleOut.size());
437         std::move(shuffleOut.begin(), shuffleOut.end(), reduceIn.begin());
438         // Initialize the reducedata struct
439         ReduceData reducedata = {.mapReduce = mapReduce, .threadNum = 0,
440             .numOfItems = reduceIn.size(), .nextToRead = 0, .logofs = ofs};
441 
442         // Measure reduce time
443         t1 = std::chrono::high_resolution_clock::now();
444         // Reduce!
445         my_pthread_mutex_lock(&mutex_er_list);
446         for (int i = 0; i < multiThreadLevel; ++i)
447         {
448             my_pthread_mutex_lock(&mutex_log_write);
449             ofs << MSG_THREAD_CREATED(THREAD_EXECREDUCE) << getTimeStr()
450                 << endl;
451             my_pthread_mutex_unlock(&mutex_log_write);
452             reducedata.threadNum = i;
453             my_pthread_create(&threadList[i], nullptr,
454                     &ExecReduce, &reducedata);
455         }
456         for (int i = 0; i < multiThreadLevel; ++i)
457         {
458             reduceResultLists[i] = std::make_pair(threadList[i],
459                     new(std::nothrow) K3_V3_LIST);
460             if (reduceResultLists[i].second == nullptr)
461             {
462                 handleError("new operator");
463             }
464         }
465         my_pthread_mutex_unlock(&mutex_er_list);
466         // Wait for reduce to finish
467         for (int i = 0; i < multiThreadLevel; ++i)
```

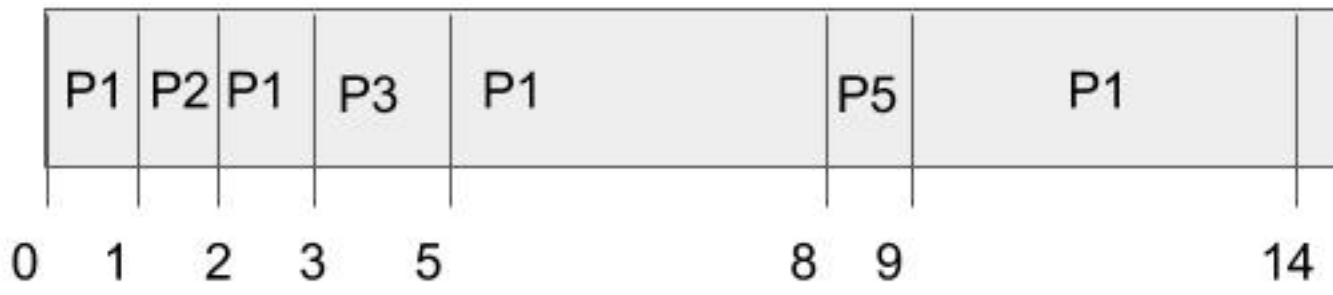
```
468     {
469         my_pthread_join(threadList[i], nullptr);
470     }
471     // "Shuffle" reduce's output!
472     finalize_output();
473     // Convert to output type and return.
474     OUT_ITEMS_LIST finalOutput(finalOutputMap.size());
475     std::move(finalOutputMap.begin(), finalOutputMap.end(),
476             finalOutput.begin());
477     // Clean up
478     cleanup();
479 
480     // Stop measuring time, write to log and return
481     t2 = std::chrono::high_resolution_clock::now();
482     auto reduceDuration = std::chrono::duration_cast
483         <std::chrono::nanoseconds>(t2 - t1).count();
484     ofs << MSG_MAP_SHUFFLE_TIME(mapShuffleDuration) << endl;
485     ofs << MSG_REDUCE_TIME(reduceDuration) << endl;
486     ofs << MSG_FRAMEWORK_END << endl;
487     return finalOutput;
488 }
489 
490 
491 /**
492  * Add <k2*,v2*> to the map-threads data structures
493  */
494 void Emit2(k2Base *key, v2Base *val)
495 {
496     // Figure out which thread is handling this function
497     pthread_t currThread;
498     currThread = pthread_self();
499     size_t end = mapResultLists.size();
500     // Add the given key,val pair to the correct list:
501     for (size_t i = 0; i < end; ++i)
502     {
503         if (pthread_equal(mapResultLists[i].first, currThread))
504         {
505             my_pthread_mutex_lock(&mutex_map_write[i]);
506             mapResultLists[i].second->push_back(
507                     std::make_pair(key, val) );
508             my_pthread_mutex_unlock(&mutex_map_write[i]);
509             return;
510         }
511     }
512 }
513 
514 /**
515  * Add <k3*,v3*> to the reduce-threads data structures
516  */
517 void Emit3(k3Base *key, v3Base *val)
518 {
519     pthread_t currThread;
520     currThread = pthread_self();
521     size_t end = reduceResultLists.size();
522     for (size_t i = 0; i < end; ++i)
523     {
524         if (pthread_equal(reduceResultLists[i].first, currThread))
525         {
526             reduceResultLists[i].second->push(
527                             std::make_pair(key, val));
528             return;
529         }
530     }
531 }
```

6 Priority.jpg

7 RR.jpg



8 SRTF.jpg



9 Search.cpp

```cpp
#include "MapReduceClient.h"
#include "MapReduceFramework.h"

#include <cerrno>
#include <string>
#include <vector>
#include <iostream>
#include <dirent.h>
#include <cstring>

#define ARG_STR_TO_FIND 1
#define ARG_DIR_START 2

#define EXIT_SUCC 0
#define EXIT_FAIL 1
#define THREAD_LEVEL 5
#define CURR_DIR "."
#define PARENT_DIR ".."

using std::string;

// Constants:
const int MIN_ARGS = 2;
const string USAGE_ERROR_MESSAGE = "Usage: <substring to search> "
    "<folders, separated by spaces>";


/**
 * Handle errors as required
 */
static void handleError(const string funcName)
{
    std::cerr << "MapReduceFramework Failure: " << funcName << " failed."
         << std::endl;
    exit(EXIT_FAIL);
}

/**
 * The directory name key (k1Derived)
 */
class DirNameKey : public k1Base
{
public:
    DirNameKey(const string d)  : dirName(d) {}

    virtual ~DirNameKey() {}

    virtual bool operator<(const k1Base &other) const override
    {
        const DirNameKey &dirother =
            dynamic_cast<const DirNameKey&> (other);
        return dirName < dirother.dirName;
    }

    const string getDirName() const
    {
        return dirName;
    }
private:
```

```
60      const string dirName;
61  };
62
63  /**
64   * The string to search in the file names
65   */
66  class StringToFind : public v1Base
67  {
68  public:
69      StringToFind(const string toFind) : str(toFind) {}
70
71      string getStr() const
72      {
73          return str;
74      }
75  private:
76      const string str;
77  };
78
79
80  /**
81   * The k2 class (k2Derived) - the filename (inside the input directories)
82   */
83  class FileNameKey1 : public k2Base
84  {
85  public:
86      FileNameKey1(const string f) : fileName(f) {}
87
88      virtual ~FileNameKey1(){}
89
90      virtual bool operator<(const k2Base &other) const override
91      {
92          const FileNameKey1 &fileOther =
93              dynamic_cast<const FileNameKey1&>(other);
94          return fileName < fileOther.fileName;
95      }
96
97      const string getFileName() const
98      {
99          return fileName;
100     }
101
102 private:
103     const string fileName;
104 };
105
106
107 /**
108  * The v2 class (v2Derived) - holds a pointer to the k2 for later
109  * deletion
110  */
111 class v2Deriv : public v2Base
112 {
113 public:
114     v2Deriv(FileNameKey1 *k2p) : k2Pointer(k2p) {}
115
116     virtual ~v2Deriv()
117     {
118         delete k2Pointer;
119         k2Pointer = nullptr;
120     }
121 private:
122     FileNameKey1 *k2Pointer;
123 };
124
125
126
127 /**
```

```
128  * The k3 class (k3Derived) - the filename (inside the input directories)
129  * This is the same as k2
130  */
131 class FileNameKey2 : public k3Base
132 {
133 public:
134     FileNameKey2(const string f) : fileName(f) {}
135 
136     virtual ~FileNameKey2(){}
137 
138     virtual bool operator<(const k3Base &other) const override
139     {
140         const FileNameKey2 &fileOther =
141             dynamic_cast<const FileNameKey2&>(other);
142         return fileName < fileOther.fileName;
143     }
144 
145     const string getFileName() const
146     {
147         return fileName;
148     }
149 
150 private:
151     const string fileName;
152 };
153 
154 /**
155  * The v3 class (v3Derived) - Holds the number of times the k3 filename
156  * appeared (in total)
157  */
158 class FileCountValue : public v3Base
159 {
160 public:
161 
162     FileCountValue(int count) : myCount(count) {}
163 
164     virtual ~FileCountValue() {}
165 
166     int getCount() const
167     {
168         return myCount;
169     }
170 
171 private:
172     const int myCount;
173 };
174 
175 /**
176  * The map and reduce functions implementation
177  */
178 class MyMapReduce : public MapReduceBase {
179 public:
180 
181     virtual void Map(const k1Base *const key, const v1Base *const val)
182         const override
183     {
184         // Downcast to k1*
185         const DirNameKey* const pdirkey = (const DirNameKey* const)key;
186         const StringToFind* const ptoFind =
187             (const StringToFind* const)val;
188 
189         // Open the directory and iterate over files in it
190         string dirName = pdirkey->getDirName();
191 
192         DIR *pdir = opendir(dirName.c_str());
193         if (pdir == nullptr)
194         {
195                 // Skip this file if it is not a directory.
```

```
196             if (errno == ENOTDIR || errno == ENOENT)
197             {
198                 return;
199             }
200             handleError("opendir");
201         }
202         struct dirent *pent = nullptr;
203         string currFile;
204         while ( (pent = readdir(pdir)) )
205         {
206             currFile = pent->d_name;
207             // Ignore '.' and '..' folders
208             if (currFile.compare(CURR_DIR) == 0 ||
209             currFile.compare(PARENT_DIR) == 0 ||
210                 // or if searched string was not found.
211             currFile.find(ptoFind->getStr()) == string::npos)
212             {
213                 continue;
214             }
215             FileNameKey1* pKey =
216                 new(std::nothrow) FileNameKey1(pent->d_name);
217             v2Deriv* pVal = new(std::nothrow) v2Deriv(pKey);
218             if (pKey == nullptr || pVal == nullptr)
219             {
220                 handleError("new");
221             }
222             Emit2(pKey, pVal);
223         }
224         if (closedir(pdir) < 0)
225         {
226             handleError("closedir");
227         }
228
229     }
230
231     virtual void Reduce(const k2Base *const key, const V2_LIST &vals)
232         const override
233     {
234         const FileNameKey1* const pfileName =
235             (const FileNameKey1* const)key;
236
237         FileNameKey2 *pKey =
238            new(std::nothrow) FileNameKey2(pfileName->getFileName());
239         FileCountValue *pVal =
240             new(std::nothrow) FileCountValue(vals.size());
241         if (pKey == nullptr || pVal == nullptr)
242         {
243             handleError("new");
244         }
245         Emit3(pKey, pVal);
246         // Delete the k2,v2 pairs
247         v2Deriv* currV2Item;
248         for (auto it = vals.begin(); it != vals.end(); ++it)
249         {
250             currV2Item = dynamic_cast<v2Deriv*>(*it);
251             delete currV2Item;
252             currV2Item = nullptr;
253         }
254     }
255 };
256
257
258 /**
259  * Delete allocated <k1,v1> <k3,v3> pairs from input/output lists
260  */
261 void cleanup(IN_ITEMS_LIST& inItems, OUT_ITEMS_LIST& outItems)
262 {
263     for (auto it = inItems.begin(); it != inItems.end(); ++it)
```

```
264     {
265         delete it->first;
266         delete it->second;
267     }
268     for (auto it = outItems.begin(); it != outItems.end(); ++it)
269     {
270         delete it->first;
271         delete it->second;
272     }
273 }
274 
275 /**
276  * Runs the framework and prints the results
277  */
278 void runMapReduce(int numOfDirs, string strToFind,
279         std::vector<string>& dirName)
280 {
281     // Prepare the input
282     IN_ITEMS_LIST inItems;
283     MyMapReduce mapReduce;
284     std::pair<k1Base*, v1Base*> currPair;
285     for (int i = 0; i < numOfDirs; ++i)
286     {
287         currPair = std::make_pair(new DirNameKey(dirName[i]),
288                 new StringToFind(strToFind));
289         inItems.push_back(currPair);
290     }
291 
292     // Run MapReduce
293     OUT_ITEMS_LIST outItems =
294         runMapReduceFramework(mapReduce, inItems, THREAD_LEVEL);
295 
296     // Print the [sorted] output
297     for (auto it = outItems.begin(); it != outItems.end(); ++it)
298     {
299         FileNameKey2* filename =
300             dynamic_cast<FileNameKey2*>(it->first);
301         FileCountValue* count =
302             dynamic_cast<FileCountValue*>(it->second);
303 
304         // For each occurence of the filename, print it.
305         for (int i = 0; i < count->getCount(); ++i)
306         {
307             std::cout << filename->getFileName() << std::endl;
308         }
309     }
310     cleanup(inItems, outItems);
311 }
312 
313 int main(int argc, char* argv[])
314 {
315     if (argc >= MIN_ARGS)
316     {
317         string strToFind = argv[ARG_STR_TO_FIND];
318         std::vector<string> dirNames(argv + ARG_DIR_START, argv + argc);
319         // Minus 2 for this file name (Search.cpp) and the str to find
320         runMapReduce(argc - 2, strToFind, dirNames);
321     }
322     else
323     {
324         std::cerr << USAGE_ERROR_MESSAGE << std::endl;
325         std::exit(EXIT_FAIL);
326     }
327     return EXIT_SUCC;
328 }
```

10 error handle.h

```
1  #ifndef _ERROR_HANDLE_H
2  #define _ERROR_HANDLE_H
3
4  #include <iostream>
5  #include <string>
6
7  #define EXIT_ERROR 1
8
9  using std::string;
10 using std::cerr;
11 using std::endl;
12
13 /**
14  * Handles an error with a function named <funcName> as specified in the
15  * exercise instructions.
16  */
17 void handleError(const string funcName)
18 {
19     cerr << "MapReduceFramework Failure: " << funcName << " failed."
20          << endl;
21     exit(EXIT_ERROR);
22 }
23
24 #endif
```

11 my pthread.h

```c
/**
 * This header is a wrapper for some of the pthread library functions
 * which adds error handling, as specified in the exercise instructions.
 * All functions don't return any value (void functions) and upon errors they
 * print the default error message and exit the process.
 */
#ifndef _MY_PTHREAD_H
#define _MY_PTHREAD_H

#include <pthread.h>
#include "error_handle.h"

#define PTHREAD_SUCCESS 0

void my_pthread_create(pthread_t *thread, const pthread_attr_t *attr,
        void *(*start_routine) (void *), void *arg)
{
    int ret_code = PTHREAD_SUCCESS;
    ret_code = pthread_create(thread, attr, start_routine, arg);
    if (ret_code != PTHREAD_SUCCESS)
    {
        handleError("pthread_create");
    }
}

void my_pthread_join(pthread_t thread, void **retval)
{
    int ret_code = PTHREAD_SUCCESS;
    ret_code = pthread_join(thread, retval);
    if (ret_code != PTHREAD_SUCCESS)
    {
        handleError("pthread_join");
    }
}

void my_pthread_mutex_init(pthread_mutex_t *mutex,
    const pthread_mutexattr_t *attr)
{
    int ret_code = PTHREAD_SUCCESS;
    ret_code = pthread_mutex_init(mutex, attr);
    if (ret_code != PTHREAD_SUCCESS)
    {
        handleError("pthread_mutex_init");
    }
}

void my_pthread_mutex_destroy(pthread_mutex_t *mutex)
{
    int ret_code = PTHREAD_SUCCESS;
    ret_code = pthread_mutex_destroy(mutex);
    if (ret_code != PTHREAD_SUCCESS)
    {
        handleError("pthread_mutex_destroy");
    }
}

void my_pthread_mutex_lock(pthread_mutex_t *mutex)
{
    int ret_code = PTHREAD_SUCCESS;
```

```
60      ret_code = pthread_mutex_lock(mutex);
61      if (ret_code != PTHREAD_SUCCESS)
62      {
63          handleError("pthread_mutex_lock");
64      }
65  }
66
67  void my_pthread_mutex_unlock(pthread_mutex_t *mutex)
68  {
69      int ret_code = PTHREAD_SUCCESS;
70      ret_code = pthread_mutex_unlock(mutex);
71      if (ret_code != PTHREAD_SUCCESS)
72      {
73          handleError("pthread_mutex_unlock");
74      }
75  }
76
77  void my_pthread_cond_init(pthread_cond_t *cond,
78          const pthread_condattr_t *attr)
79  {
80      int ret_code = PTHREAD_SUCCESS;
81      ret_code = pthread_cond_init(cond, attr);
82      if (ret_code != PTHREAD_SUCCESS)
83      {
84          handleError("pthread_cond_init");
85      }
86  }
87
88  void my_pthread_cond_destroy(pthread_cond_t *cond)
89  {
90      int ret_code = PTHREAD_SUCCESS;
91      ret_code = pthread_cond_destroy(cond);
92      if (ret_code != PTHREAD_SUCCESS)
93      {
94          handleError("pthread_cond_destroy");
95      }
96  }
97
98  void my_pthread_cond_signal(pthread_cond_t *cond)
99  {
100     int ret_code = PTHREAD_SUCCESS;
101     ret_code = pthread_cond_signal(cond);
102     if (ret_code != PTHREAD_SUCCESS)
103     {
104         handleError("pthread_cond_signal");
105     }
106 }
107
108 void my_pthread_cond_wait(pthread_cond_t *cond, pthread_mutex_t *mutex)
109 {
110     int ret_code = PTHREAD_SUCCESS;
111     ret_code = pthread_cond_wait(cond, mutex);
112     if (ret_code != PTHREAD_SUCCESS)
113     {
114         handleError("pthread_cond_wait");
115     }
116
117 }
118
119 void my_pthread_cond_timedwait(pthread_cond_t *cond,
120         pthread_mutex_t *mutex, const struct timespec *abstime)
121 {
122     /* int ret_code = PTHREAD_SUCCESS;
123     ret_code = */pthread_cond_timedwait(cond, mutex, abstime);
124     /* if (ret_code != PTHREAD_SUCCESS)
125     {
126         handleError("pthread_cond_timedwait");
127     } */
```

```
128
129 }
130
131 #endif
```