**Algorithms**

Programming Assignment #1

Sorting

**Submission & Online Resources: on NTU COOL**

**Introduction:**

In this PA, you are required to implement various sorters that we learnt in class. You can download the PA1.tarfile from NTU COOL website. Decompress it using Linux command,

tar -xvf PA1.tar

You can see the following directories after decompressing it.

|  |  |
| --- | --- |
| **Name** | **Description** |
| bin/ | Directory of binary file(s) |
| doc/ | Directory of documents |
| inputs/ | Directory of unsorted data |
| lib/ | Directory of library source code |
| outputs/ | Directory of sorted data |
| src/ | Directory of source code |
| utility/ | Directory of checker |

**Input/output Files:**

In the input file (\*.in), the first two lines starting with ‘#’ are just comments. Except comments, each line contains two numbers: index followed by the unsorted number. The range of unsorted number is between 0 and 1,000,000. Two numbers are separated by a space. For example, the file 5.case1.in contains five numbers

|  |
| --- |
| # 5 data points  # index number  0 16  1 13  2 0  3 6  4 7 |

The output file(\*.out) is actually the same as the input file except that the numbers are sorted in *increasing* order. For example, 5.case1.out is like:

|  |
| --- |
| # 5 data points  # index number  0 0  1 6  2 7  3 13  4 16 |

**PLOT:**

You can visualize your unsorted/sorted numbers by using the gnuplot tool by the command gnuplot. After that, please key in the following

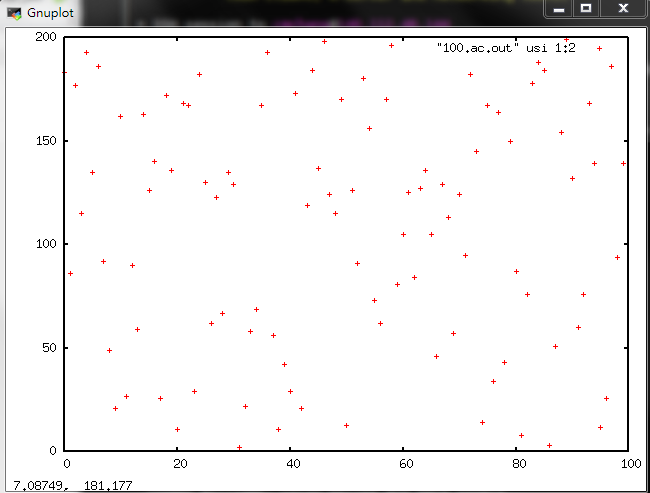
|  |
| --- |
| set xrange [0:5]  set yrange [0:20]  plot "5.case1.in" usi 1:2  plot "5.case1.out" usi 1:2  # if you want to save to png files  set terminal png  set output "5.case1.out.png"  replot |

You need to allow X-window display to see the window if you are login remotely. For more gnuplot information, see

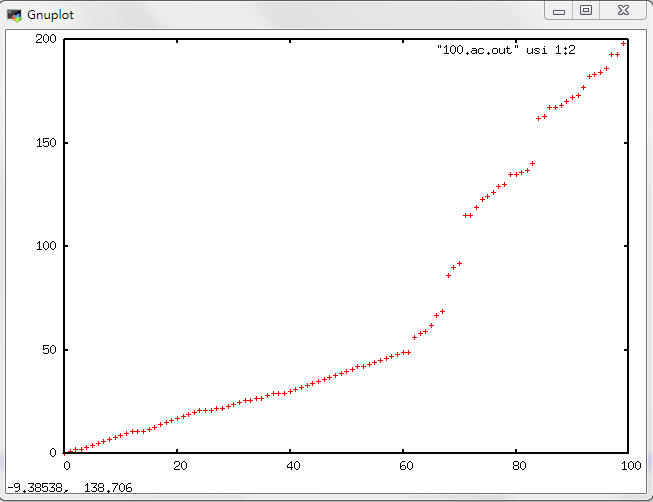
<http://people.duke.edu/~hpgavin/gnuplot.html>

Here are the “*Before*” and “*After*” sorting plots for a benchmark with 100 numbers.

*Before* sort：



*After* sort：



**Command line parameters:**

In the command line, you are required to follow this format:

NTU\_sort –[IS|MS|QS|HS] <input\_file\_name> <output\_file\_name>

where IS represents insertion sort, MS is merge sort, QS is quick sort and HS is heap sort. The square bracket with vertical bar ‘[IS|MS|QS|HS]’ means that only one of the four versions is chosen.

The angle bracket <input\_file\_name> should be replaced by the name of the input file, \*.[case1|case2|case3].in, where case1 represents test case in random order, case2 is test case in increasing order, and case3 is test case in reverse order. For the best case, all the numbers are sorted in increasing order. For the worst case, all numbers are sorted in descending order. For the average case, numbers are in random order.

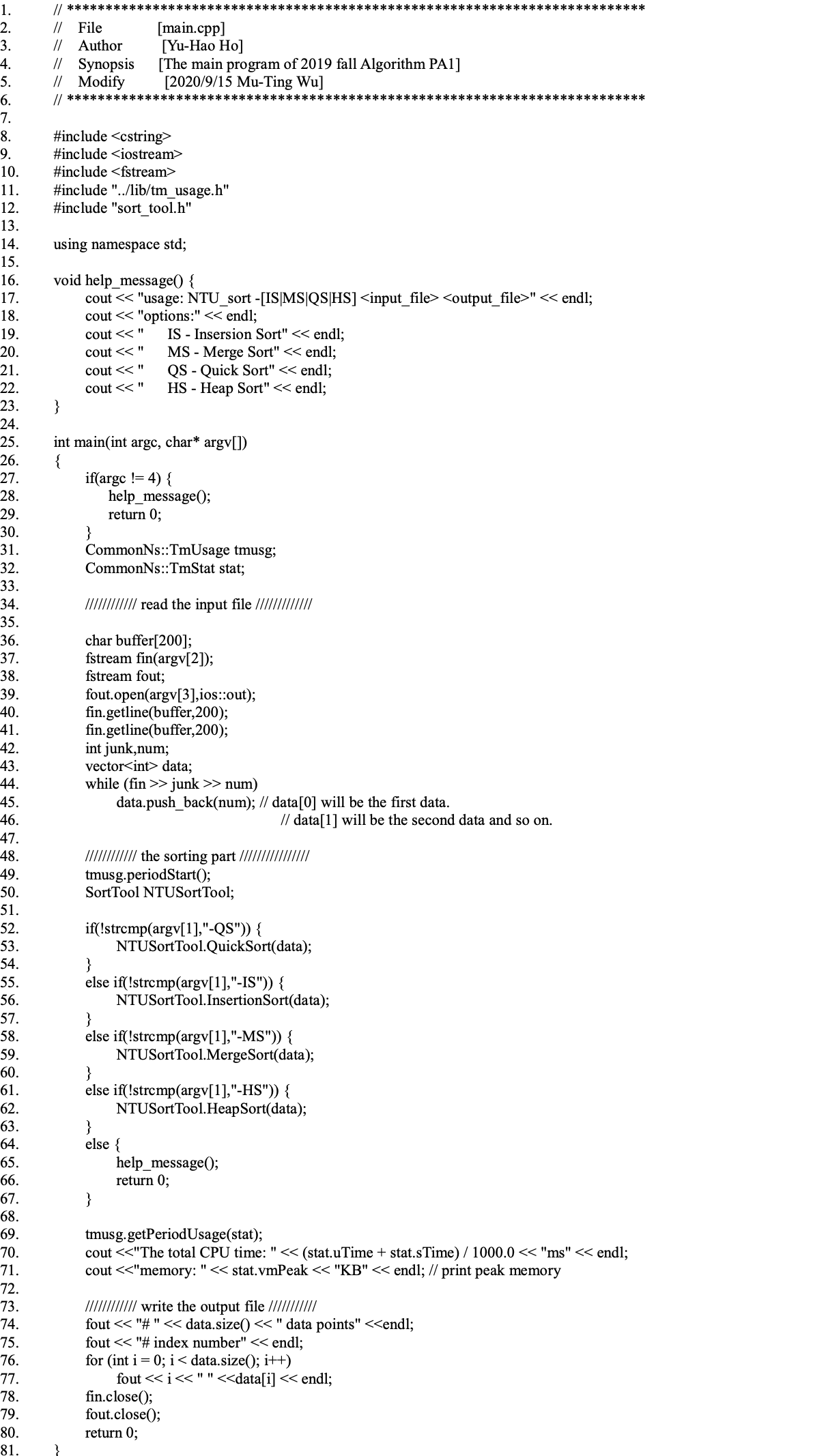
The output file names are \*.[case1|case2|case3].out. Please note that you do NOT need to add ‘[|]’ or ‘<>’ in your command line. For example, the following command sorts 1000.case1.inand outputs to 1000.case1.outusing insertion sort.

./bin/NTU\_sort -IS inputs/1000.case1.in outputs/1000.case1.out

**Source code files:**

**Please notice that all of the source code files have been already finished except sort\_tool.cpp*.*** You only need to complete the different sorting functions of class SortTool insort\_tool.cpp. You can still modify other source code files if you think it is necessary. The following will simply introduce the source code files.

main.cpp: main program for PA1



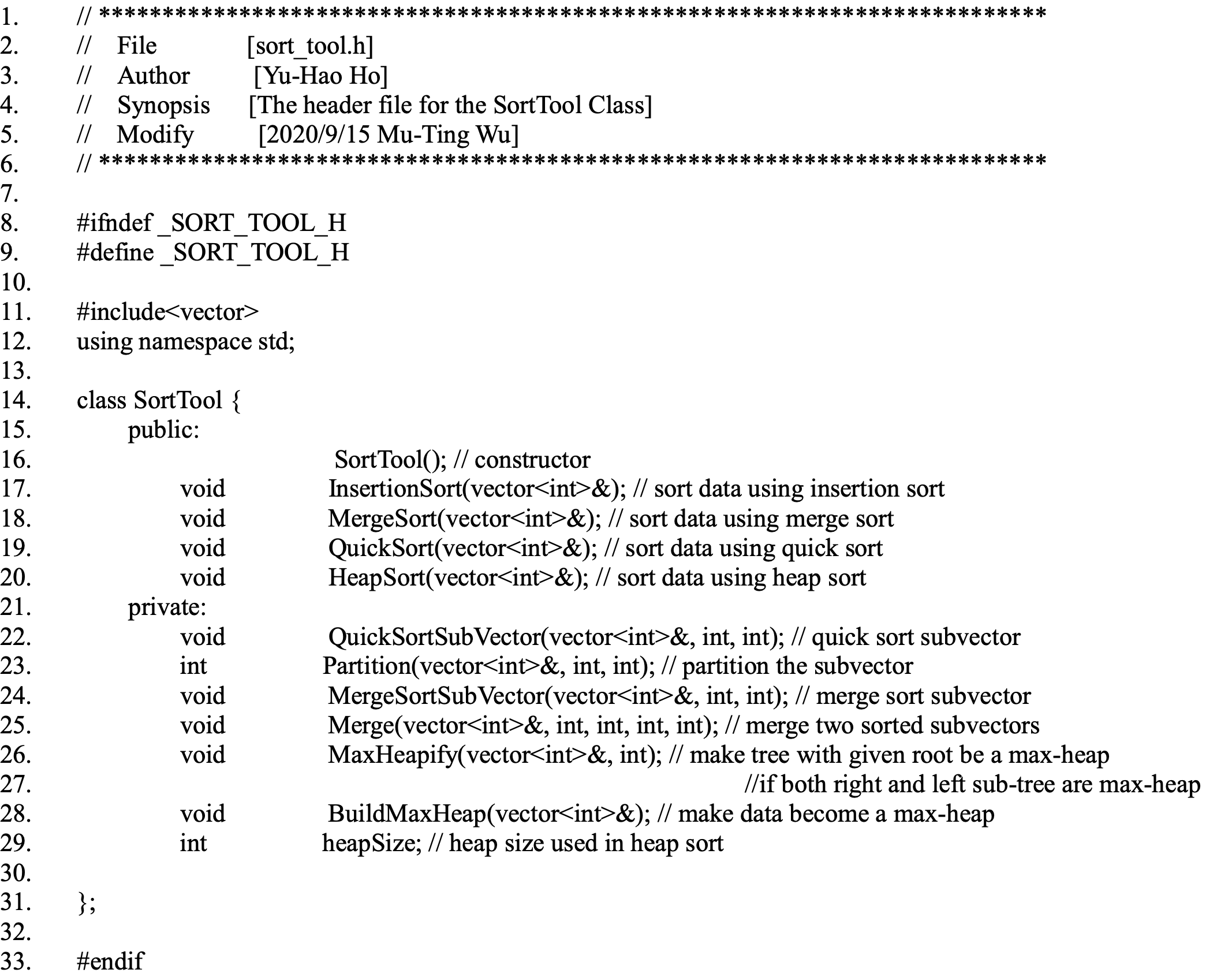
**main.cpp**

Lines 36-46: parse unsorted data from input file and push them into the vector.

Lines 52-67: call different function depending on given command.

Lines 74-77: write the sorted data file.

sort\_tool.h: the header file for the SortTool Class



**sort\_tool.h**

Lines 17-20: sort function which will be called in main.cpp*.*

Line 22: This function will be used in quick sort. It will sort sub vector with given lower and upper bound. This function should be implemented to partition the sub vector and recursively call itself.

Line 23: This function will be used in quick sort and should be implemented to partition the sub vector.

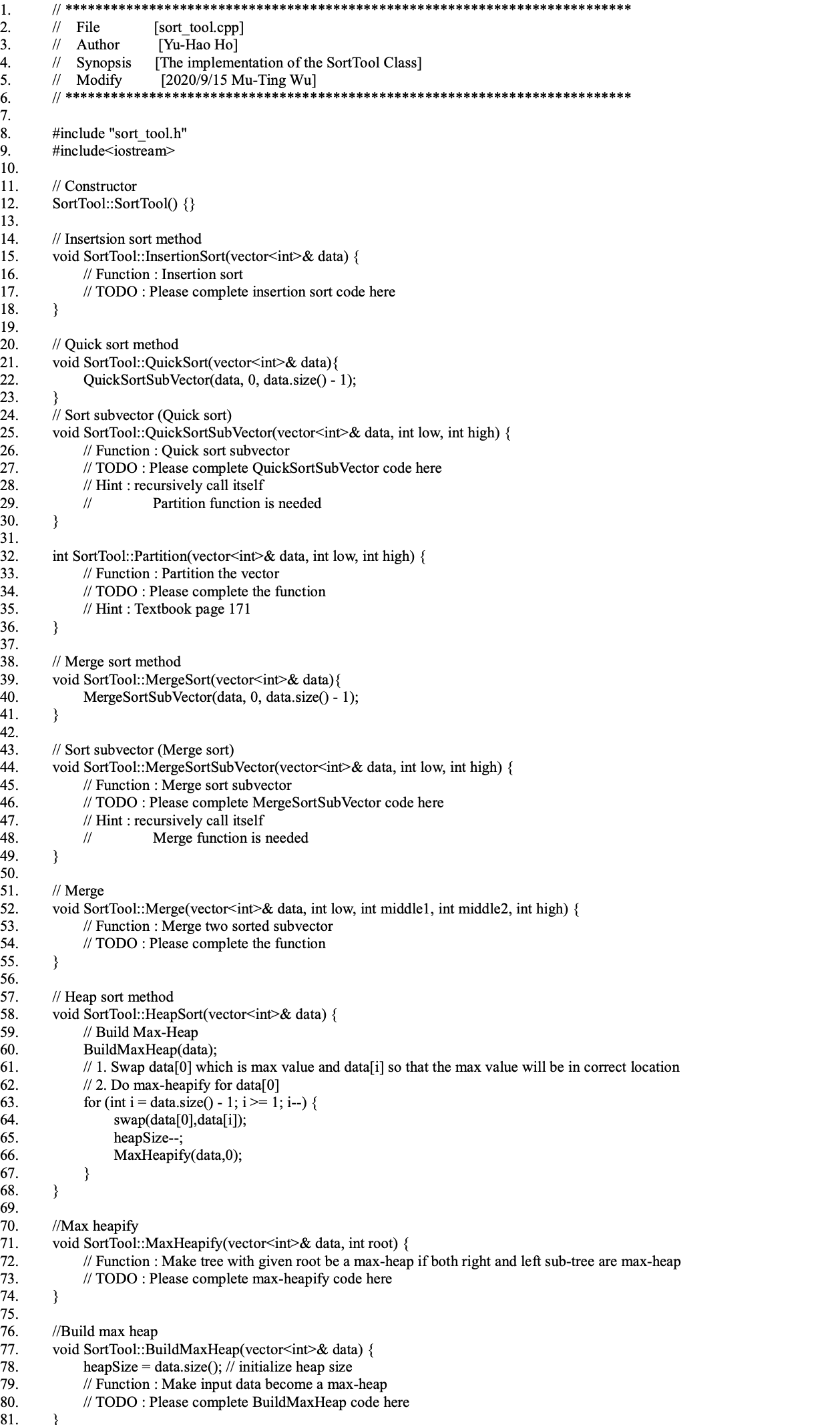
Line 24: This function will be used in merge sort. It will sort sub vector with given lower and upper bound. This function should be implemented to call itself for splitting and merging the sub vector.

Line 25: This function will be used in merge sort and should be implemented to merge two sorted sub vectors.

Line 26: This function will be used in heap sort and should be implemented to make the tree with given root be a max-heap if both of its right subtree and left subtree are max-heap.

Line 28: This function will be used in heap sort and should be implemented to make input data be a max-heap.

sort\_tool.cpp: the implementation of the SortTool Class



**sort\_tool.cpp**

Lines 15-18: please complete the function of insertion sort here*.*

Lines 21-23: the function of quick sort will call function of Sorting sub-vector and give initial lower/upper bound.

Lines 25-30: please complete the function of sorting sub-vector using quick sort algorithm here.

Lines 32-36: please complete the function of partition here.

Lines 39-41: the function of merge sort will call function of Sorting sub-vector and give initial lower/upper bound.

Lines 44-49: please complete the function of sorting sub-vector using merge sort algorithm here.

Lines 52-55: please complete the function of merging two sorted sub-vector here.

Lines 58-68: the function of heap sort will build max-heap first. And then, exchange data iteratively.

Lines 71-74: please complete the function of max-heapify which makes the tree with given root be a max-heap if its right and left sub-tree are both max-heap.

Lines 77-81: please complete the function of building max-heap with given input data.

**Requirements:**

1. Please check the source code files under the src directory. You may need to complete the functions of class SortTool in sort\_tool.cpp. You can also modify main.cpp and sort\_tool.h if you think it is necessary.

2. Your source code must be written in C or C++. The code must be executable on EDA Union Lab machines.

3. In your report, compare the running time of four versions of different input sizes. Please fill in the following table. Please use -O2 optimization and turn off all debugging message.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Input size | IS | | MS | | QS | | HS | |
|  | CPU time  (s) | Memory  (KB) | CPU time  (s) | Memory  (KB) | CPU time  (s) | Memory  (KB) | CPU time  (s) | Memory  (KB) |
| 4000.case1 |  |  |  |  |  |  |  |  |
| 4000.case2 |  |  |  |  |  |  |  |  |
| 4000.case3 |  |  |  |  |  |  |  |  |
| 16000.case1 |  |  |  |  |  |  |  |  |
| 16000.case2 |  |  |  |  |  |  |  |  |
| 16000.case3 |  |  |  |  |  |  |  |  |
| 32000.case1 |  |  |  |  |  |  |  |  |
| 32000.case2 |  |  |  |  |  |  |  |  |
| 32000.case3 |  |  |  |  |  |  |  |  |
| 1000000.case1 |  |  |  |  |  |  |  |  |
| 1000000.case2 |  |  |  |  |  |  |  |  |
| 1000000.case3 |  |  |  |  |  |  |  |  |

4. In your report, draw figures to show the growth of run time as a function of input size and try to analyze the slopes of the curves as well as their relation (as the following example, where each curve represents an algorithm.) **Please note that you should transfer the run time and input size to log scale first, then draw the figures.** Your figures should be clear and easy to distinguish the curves, where **the slope of each curve should be specified in your report**.

|  |  |  |
| --- | --- | --- |
| page8image5874000.png | page8image5878368.png | page8image5874624.png |

You can skip the test case if the run time is more than 3 minutes.

5. Notice: You are not allowed to include the header <algorithm> or <queue> in STL!

**Compile**

We expect your code can compile and run in this way.

Type the following commands under <student\_id>\_pa1 directory,

make

cd bin

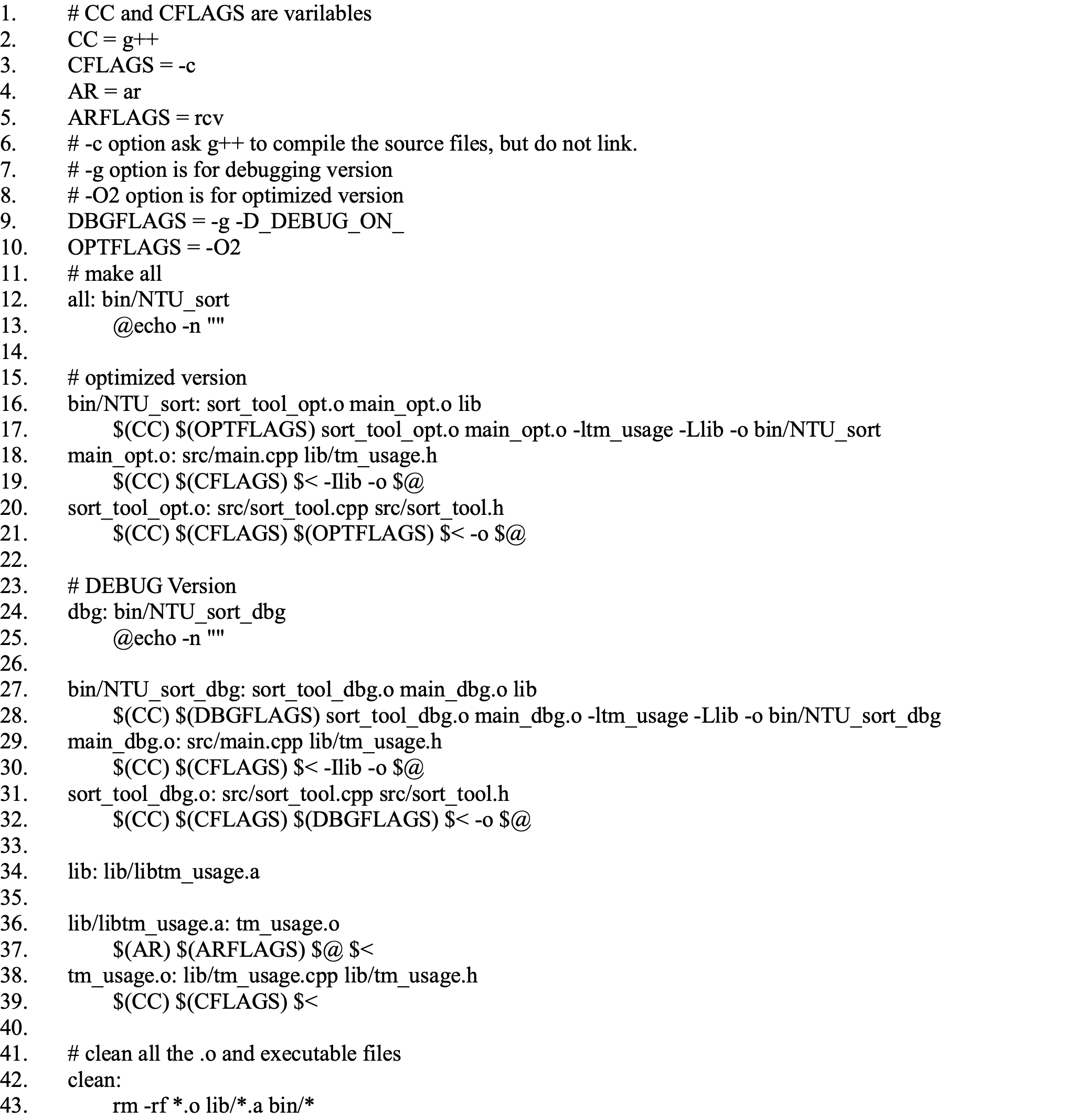
./NTU\_sort –[IS|MS|QS|HS] <input\_file\_name> <output\_file\_name>

We provide the sample makefile, **please modify into yours if needed**.

**Control the stack size**

To prevent stack overflow cause by the recursion function calls, please set the stack size to 256MB using the following Linux comment:

ulimit -s 262144



**m****a****k****e****f****i****l****e**

Lines 38-39: compile the object file tm\_usage.o from tm\_usage.cppand tm\_usage.h

Lines 36-37: archive tm\_usage.o into a static library file libtm\_usage.a. Please note that library must start with lib and ends with .a.

Line 37: this small library has only one objet file. In a big library, more than one objective files can be archived into a single lib\*.afile like this

ar rcv libx.a file1.o [file2.o …]

Lines 12-21: When we type ‘make’ without any option the makefile will do the first command (line.12 in this sample). Thus, we can compile the optimization version when we type ‘make’. This version invokes options ‘-O2’ for speed improvement. Also ‘\_DEBUG\_ON\_’ is not defined to disable the printing of arrays in sort\_tool.cpp.

Lines 23-32: Compile the debug version when we type ‘make dbg’. This version invokes options ‘-g’ (for DDD debugger) and also ‘-D\_DEBUG\_ON\_’ to enable the printing of arrays in sort\_tool.cpp.

Lines 13, 25: @echo –n “” will print out the message in “”. In this sample we print nothing.

Notice: $< represent the first dependency.

$@ represent the target itself.

Example: a.o : b.cpp b.h

$(CC) $(CFLAGS) $(DBGFLAGS) $< -o $@

$< = b.cpp $@ = a.o

You can find useful information here.

[Makefile Tutorial By Example](https://makefiletutorial.com/)

**Validation:**

You can verify your answer very easily by comparing your output with case2 which is the sorted input. Or you can see the gnuplot and see if there is any dot that is not sorted in order.

Also, you can use our result checker which is under utility directory to check whether your result is correct or not. To use this checker, simply type

./PA1\_Result\_Checker <input\_file> <your\_output\_file>

Please notice that it will not check whether the format of result file is correct or not. You have to check the format by yourself if you modify the part of writing output file in main.cpp.

**Submission:**

You need to create a directory named **<student\_id>\_pa1/** (e.g. b09901000\_pa1/) (**student id should start with a lowercase letter**) which must contain the following materials:

1. A directory named **src/** contains your source codes: only \*.h, \*.hpp, \*.c, \*.cpp are allowed in src/, and no directories are allowed in src/;
2. A directory named **bin/** containing your executable binary named **NTU\_sort**;
3. A directory named **doc/** containing your report;
4. A makefile named **makefile** that produces an executable binary from your source codes by simply typing “make”: the binary should be generated under the directory <student\_id>\_pa1/bin/;
5. A text readme file named **README** describing how to compile and run your program;
6. A report named **report.pdf** on the data structures used in your program and your findings in this programming assignment.

We will use our own test cases, so do NOT include the input files.

In summary, you should at least have the following items in your \*.tgz file.

src/<all your source code>

lib/<library file>

bin/NTU\_sort

doc/report.pdf

makefile

README

The submission filename should be compressed in a single file <student\_id>\_pa1.tgz. (e.g. b09901000\_pa1.tgz). You can use the following command to compress a whole directory:

tar -zcvf <filename>.tgz <dir>

For example, go to the same level as PA1 directory, and type

tar -zcvf b09901000\_pa1.tgz b09901000\_pa1/

Please submit a single *\*.tgz* file to NTU COOL system before **3/20 (Sun.) 13:00**.

You are required to run the checksubmitPA1 script to check if your .tgz submission file is correct. Suppose you are in the same level as PA1 directory

bash ./PA1/utility/checkSubmitPA1.sh b09901000\_pa1.tgz

Please note the path must be correct. If you are located in the ~/ directory, then

‘./PA1/utility/checkSubmitPA1.sh’ means the path

~/PA1/utility/checkSubmitPA1.sh

and b09901000\_pa.tgz means the path ~/b99901000\_pa1.tgz

Your program will be graded by automatic grading script. Any mistake in the submission will cost at least 20% penalty of your score. Please be very careful in your submission.

**Grading:**

70% correctness (including submission correctness and implementation correctness)

20% file format and location

10% report

**NOTE:**

1. TA will check your source code carefully. Copying other source codes can result in zero grade for all students involved.
2. Implementation correctness means to follow the guideline on the handout to write the codes. Wrong implementation will result in penalty even if the output is correct.

**FAQ:**

**Q0: Who can answer my questions regarding PA#1?**

A0: Please send an email to TA, Nick Liang, at d10943012@ntu.edu.tw.

**Q1: Where is my EDA Union account and password?**

A1: NTU COOL > 成績 > Account & Password > 評語.

**Q2: How to log in to the server?**

A2: Please see the file “Recitation\_0217\_Rules & Tutorial.pdf” in NTU COOL > 課程內容 > Recitation.

Someone cannot log in by

“ssh algs22xxx@edaunion.ee.ntu.edu.tw -p <port>”

can try another command

“ssh -p <port> alg22sxxx@edaunion.ee.ntu.edu.tw”.

**Q3: Should we take care of the error handling issue of the testbench?**

A3: The testbench for this PA is correct. However, it is always better to have your error handling protection.

**Q4: What is the runtime limit for each case?**

A4: The runtime limit for each case is 3 minutes.

**Q5: Can I use C++11? Can I use Mac OS? Can I use …?**

A5: Yes, you can. However, we will use and only use the EDA Union server for evaluation. Hence, whatever platform/setting you use, please test your code on the EDA Union server before submitting your final version. The submission that fails to compile and/or run on the EDA Union server will be severely penalized.

**Q6: What is my EDA Union account?**

A6: Your EDA Union account (a.k.a Web ID) is “alg21XXX” starts with the lower letter “a”.

**Q7: Runtime warning message: cannot get memory usage**

A7: Our code does not support checking memory usage on Mac OS.  To check memory usage, compile and run the code on the EDA Union server.

**Q8: Quicksort runtime error message: segmentation fault**

A8: If the code is correct, this might be caused by stack overflow.  Also, other memory-related issues might be reported in the error message.  Enlarging the stack size would help.  See section “Control the stack size” in the pdf.

**Q9: Do the submission files include the lib/ folder?**

A9: Yes.  You are also allowed to modify the codes in lib/, so it’s recommended that you attach them in your submission.

**Q10: Permission denied when using PA1 checker.**

A10: Type “chmod u+x ./utility/PA1\_result\_checker” and run again.

**Q11: If there is any compile error (e.g. “Error: expecting string instruction after `rep'”)**

A11: “Google” it first or try to compile it on another server.