#include <stdio.h>

#include <stdlib.h>

#include <string.h>

#include <limits.h>

#include <time.h>

#define MAX\_FILENAME\_LENGTH 255

void mergeSort(int arr[], int l, int r);

void merge(int arr[], int l, int m, int r);

void externalSort(char\* inputFileName, char\* outputFileName, int bufferSize, int maxTemp);

void mergeSortFiles(char\* outputFile, char\*\* tempFiles, int numFiles);

void generateRandomInputFile(char\* filename, int numInts);

void doWork(int buf, int nums, char\* inputFile, char\* outputFile, int maxTemp);

**int main()**

{

printf("External sort\n");

int progRuns = 5;

for (int i=0; i< progRuns; i++)

{

if (i == 0)

{

char\* inputFileName = "data5k.txt";

char\* outputFileName = "sorted5k.txt";

int bufferSize = 100;

int numInts = 5000;

int maxTemp = 50;

doWork(bufferSize, numInts, inputFileName, outputFileName, maxTemp);

} else if (i == 1)

{

char\* inputFileName = "data40k.txt";

char\* outputFileName = "sorted40k.txt";

int bufferSize = 1000;

int numInts = 40000;

int maxTemp = 40;

doWork(bufferSize, numInts, inputFileName, outputFileName, maxTemp);

} else if (i == 2)

{

char\* inputFileName = "data300k.txt";

char\* outputFileName = "sorted300k.txt";

int bufferSize = 10000;

int numInts = 300000;

int maxTemp = 30;

doWork(bufferSize, numInts, inputFileName, outputFileName, maxTemp);

} else if (i == 3)

{

char\* inputFileName = "data2M.txt";

char\* outputFileName = "sorted2M.txt";

int bufferSize = 100000;

int numInts = 2000000;

int maxTemp = 20;

doWork(bufferSize, numInts, inputFileName, outputFileName, maxTemp);

} else

{

char\* inputFileName = "data10M.txt";

char\* outputFileName = "sorted10M.txt";

int bufferSize = 200000;

int numInts = 10000000;

int maxTemp = 50;

doWork(bufferSize, numInts, inputFileName, outputFileName, maxTemp);

}

}

return EXIT\_SUCCESS;

}

//Main do work function

void doWork(int buf, int nums, char\* inputFile, char\* outputFile, int maxTemp)

{

printf(" Sorting file: %s\n", inputFile);

printf(" Output file: %s\n", outputFile);

printf(" Buffer size: %d\n", buf);

generateRandomInputFile(inputFile, nums);

clock\_t startTime = clock();

externalSort(inputFile, outputFile, buf, maxTemp);

clock\_t endTime = clock();

double elapsed = (double)(endTime - startTime) / CLOCKS\_PER\_SEC;

printf("\n\nElapsed time to sort %s using %d temp files!", inputFile, maxTemp);

printf("\n %.4f minutes", elapsed / 60);

printf("\n %.4f seconds", elapsed);

printf("\n %.4f milli-seconds\n", elapsed \* 1000);

printf("\nDone\n");

printf("...............");

}

//External Sort function

void externalSort(char\* inputFileName, char\* outputFileName, int bufferSize, int MAX\_TEMP\_FILES)

{

FILE\* inputFile = fopen(inputFileName, "r");

if (!inputFile) {

perror("Failed to open input file");

exit(EXIT\_FAILURE);

}

int numTempFiles = 0;

int\* buffer = (int\*)malloc(bufferSize \* sizeof(int));

char\*\* tempFiles = malloc(MAX\_TEMP\_FILES \* sizeof(char\*));;

char tempFileName[MAX\_FILENAME\_LENGTH];

int bufferIndex = 0, value;

while (fscanf(inputFile, "%d", &value) != EOF) {

buffer[bufferIndex++] = value;

if (bufferIndex == bufferSize) {

mergeSort(buffer, 0, bufferSize - 1);

if (numTempFiles >= MAX\_TEMP\_FILES)

{

fprintf(stderr, "Maximum number of temporary files reached. Aborting.\n");

exit(EXIT\_FAILURE);

}

snprintf(tempFileName, sizeof(tempFileName), "tempFile%d.txt", numTempFiles);

FILE\* tempFile = fopen(tempFileName, "w");

for (int i = 0; i < bufferSize; i++) {

fprintf(tempFile, "%d\n", buffer[i]);

}

fclose(tempFile);

tempFiles = realloc(tempFiles, (numTempFiles + 1) \* sizeof(char\*));

tempFiles[numTempFiles] = strdup(tempFileName);

numTempFiles++;

bufferIndex = 0;

}

}

// Handle last set of values (if any)

if (bufferIndex > 0)

{

mergeSort(buffer, 0, bufferIndex - 1);

snprintf(tempFileName, sizeof(tempFileName), "tempFile%d.txt", numTempFiles);

FILE\* tempFile = fopen(tempFileName, "w");

for (int i = 0; i < bufferIndex; i++) {

fprintf(tempFile, "%d\n", buffer[i]);

}

fclose(tempFile);

tempFiles = realloc(tempFiles, (numTempFiles + 1) \* sizeof(char\*));

tempFiles[numTempFiles] = strdup(tempFileName);

numTempFiles++;

}

fclose(inputFile);

free(buffer);

mergeSortFiles(outputFileName, tempFiles, numTempFiles);

// Cleanup

for (int i = 0; i < numTempFiles; i++) {

free(tempFiles[i]);

}

free(tempFiles);

}

// Implement mergeSort and mergeSortFiles based on the initial guidance.

// Simple Merge Sort implementation for sorting each chunk

void merge(int arr[], int l, int m, int r) {

int i, j, k;

int n1 = m - l + 1;

int n2 = r - m;

int L[n1], R[n2];

for (i = 0; i < n1; i++)

L[i] = arr[l + i];

for (j = 0; j < n2; j++)

R[j] = arr[m + 1 + j];

i = 0;

j = 0;

k = l;

while (i < n1 && j < n2) {

if (L[i] <= R[j]) {

arr[k] = L[i];

i++;

} else {

arr[k] = R[j];

j++;

}

k++;

}

while (i < n1) {

arr[k] = L[i];

i++;

k++;

}

while (j < n2) {

arr[k] = R[j];

j++;

k++;

}

}

//Merge sort

void mergeSort(int arr[], int l, int r) {

if (l < r)

{

int m = l + (r - l) / 2;

mergeSort(arr, l, m);

mergeSort(arr, m + 1, r);

merge(arr, l, m, r);

}

}

// Function to merge sorted chunks from temporary files

void mergeSortFiles(char\* outputFile, char\*\* tempFiles, int numFiles)

{

FILE\* outFile = fopen(outputFile, "w");

FILE\*\* fps = malloc(numFiles \* sizeof(FILE\*));

int\* values = malloc(numFiles \* sizeof(int));

char line[256];

int activeFiles = numFiles;

for (int i = 0; i < numFiles; i++) {

fps[i] = fopen(tempFiles[i], "r");

if (fgets(line, sizeof(line), fps[i]) != NULL) {

values[i] = atoi(line);

} else {

values[i] = INT\_MAX;

activeFiles--;

}

}

while (activeFiles > 0) {

int minIndex = 0;

int minValue = INT\_MAX;

for (int i = 0; i < numFiles; i++) {

if (values[i] < minValue) {

minValue = values[i];

minIndex = i;

}

}

if (minValue != INT\_MAX) {

fprintf(outFile, "%d\n", minValue);

if (fgets(line, sizeof(line), fps[minIndex]) != NULL) {

values[minIndex] = atoi(line);

} else {

values[minIndex] = INT\_MAX;

activeFiles--;

}

}

}

for (int i = 0; i < numFiles; i++)

{

fclose(fps[i]);

remove(tempFiles[i]); // Ensure temporary files are removed.

}

fclose(outFile);

free(fps);

free(values);

}

//Function to generate the input file having random ints

//Generate the random numbers input file (to be later sorted)

void generateRandomInputFile(char\* filename, int numInts)

{

FILE\* file = fopen(filename, "w");

if (file == NULL) {

perror("Error creating input file");

exit(EXIT\_FAILURE);

}

srand(time(NULL));

for (int i = 0; i < numInts; i++) {

fprintf(file, "%d\n", rand());

}

fclose(file);

}