Assignment 4 – Sets and Sorting

Sebastian Avila

CSE 13S - Fall 2023

Purpose

The purpose of this program is to implement a set of Set functions and a set of Sorting algorithms. These sorting algorithms are heap sort, shell sort, quick sort, insert sort, and batcher sort.

How to Use the Program

To use this program you will need to have several files in the same directory. The files are set.c\,, set.h, insert.c, insert.h, , shell.c, shell.h, heap.c, heap.h, quick.c, quick.h, batcher.c, batcher.h, sorting.c, stats.c, stats.h, and Makefile. With these files in the same directory, use command make clean to remove any preexisting object files. Then, use command make format to format all the header and source files. Then the same directory use command make to compile the program. Then run the program using ./sorting with a command line argument. The possible command line arguments and their meanings are as follows:

- -H: Display program help and usage message
- -a: Enable all sorting algorithms
- -h : Enable Heap sort
- -b : Enable Batcher sort
- -s : Enable Shell sort
- -q : Enable Quick sort
- -i : Enable Insertion sort
- -n length: Set the array size to length. If this argument is missing, the default length is 100.
- -p elements: Print out elements number of elements from the array. If this agrument is missing, the default value is 100. If the array size is less than elements the entire array will be printed out and nothing more. If elements is set to 0, nothing will be printed.
- -r seed: Set the random seed to seed. If this argument is missing, the default seed is 13371453.

The program uses several optional compiler flags:

- -Wall: This flag enables all warning messages.
- -Werror: This flag turns all warnings into errors.
- -Wextra: This flag enables extra warning flags that are not enabled by -Wall.
- -Wstrict-prototypes: This flag warns if a function is declared or defined without specifying the argument types.
- -pedantic: This flag issues all the warnings demanded by strict ISO C and ISO C++.
- -lm: This flag links the math.h library. This allows the program to access and use the functions from the math.h library.

Program Design

The program starts by initializing a set with 8 bits all set to 0. This set represents a set of sorting algorithms. It then designates each type of sorting as a certain bit. It also initializes three variable, elements, seed, size, and sets them to their default values. It then checks the command line arguments. If an incorrect argument is received, an error message and a usage message is printed. The arguments pertaining to the sort options add the corresponding bit to the set when the argument is received. The other three arguments take in a string and covert it into an integer then store that integer in a variable that was initialized before. After the command arguments are dealt with, it initializes a variable stats of type Stats, creates a variable with a hexadecimal value of 0x3fffffff. That is 32 bits, with the lower 30 bits set to 1. This variable is used to mask the random numbers generated into only the lower 30 bits. Memory is then allocated using the size of the first element of the array by the size of the array. Next, a series of if statements checks whether each sort is a member of the set. If it is, the stats variable and the array are reset using their corresponding functions. The array is then sorted using the sorting algorithm function represented by the set member that was checked, the stats are printed using a function, then the elements of the array are printed using a different function. Once all of the if statements have been checked, the program frees the memory previously allocated for the array and then returns 0.

Algorithms

All algorithms in this section were given in the assignment instructions. [1] Psuedocode for the algorithms is shown in the function description section of this report.

- Insertion Sort: This algorithm assumes the first element is already sorted. It then moves to the next element and compares it to the first. If it is less than the first element it swaps the elements. It repeats this process for the remaining elements, each time assuming the previous element is already sorted.
- Shell Sort: This is a version of insertion sort. Shell sort repeatedly sorts sub arrays that are separated by a gap that is reduced continuously. After the gap is 1, the array is completely sorted.
- Heap Sort: The first step in the heap sort algorithm is to create a max heap from the array that is to be sorted. A max heap array is a binary tree where each node is greater than or equal to its children. Once the array is transformed into a max heap, the root of the heap is swapped with the last element in the array and the heap size is reduced by 1. Next, the heap is fixed to maintain the max heap structure. The last two steps are repeated until the entire array is sorted.
- Quick Sort: Quick sort works by choosing a pivot element and partitioning the other elements into two sub arrays according to whether they are less than or greater than the pivot. The sub arrays are then sorted recursively.
- ullet Batcher Sort : Batcher sort k-sorts the even and odd sub-sequences of an array, where k is a power of 2, and then merges them.

Function Descriptions

Sorting

• void make_array(int *A, int seed, int size, int mask): This function takes 4 parameters: three integers seed, size, mask, and a pointer to an integer array A. It does not return anything. This function resets the array to the same random unsorted values that it was before being sorted.

```
void make_array(int *A, int seed, int size, int mask)
    srandom(seed)
    for i, 0 to i < size
        A[i] = random() AND mask</pre>
```

• void print_elements(int *A, int size, int elements): This function takes in three parameters: a pointer to an integer array A, an integer representing its size, and an integer elements that represents how many elements to print. It does not return anything. Its purpose is to print the elements of the array in 5 columns with each element having a width of 13.

```
void print_elements(int *A, int size, int elements)
    j = elements
    if size < elements then
        j = size
    for i, 0 to i < j
        print A[i] with width of 13
        if ((i + 1) % 5) = 0) then
            print newline
        else if i + 1 >= j then
            print new line
```

Sets

• Set set_empty(void): This function takes no parameters and returns a set. It purpose is to return an empty set. This is a set where all bits are equal to 0.

• Set set_universal(void): This function takes no parameters and returns a set. Its purpose is to return a set in which every possible number is part of the set.

```
Set set_universal(void)
Set u = 11111111 (0xff)
return u
```

• Set set_insert(Set s, int x): This function takes two parameters: a set s and an integer x. It returns a set. Its purpose is to insert the integer x into the set s and return the set.

```
Set set_insert(Set s, int x)
    s = s OR (0x01 << x)
    return s</pre>
```

• Set set_remove(Set s, int x): This function takes two parameters: a set s and an integer x. It returns a set. Its purpose is to remove the integer x from the set s and return the set.

```
Set set_remove(Set s, int x)
    s = s AND set_complement(0x01 << x)
    return s</pre>
```

• bool set_member(Set s, int x): This function takes two parameters: a set s and an integer x. It returns a bool (true or false) indicating whether the integer x is within the set s.

```
bool set_member(Set s, int x)
  if (s AND (0x01 << x) == 1 then
    return true
  return false</pre>
```

• Set set_union(Set s, Set t): This function takes two parameters of type set: s and t. It returns a set containing only the bits corresponding to members that are equal to 1 in either s or t.

```
Set set_union(Set s, Set t)|
   s = s OR t
   return s
```

• Set set_intersect(Set s, Set t)—: This function takes two parameters of type set: s and t. It returns a set containing only the bits corresponding to members that are equal to 1 in both s and t.

```
Set set_intersect(Set s, Set t)|
    s = s AND t
    return set
```

• Set set_difference(Set s, Set t): This function takes two parameters of type set: s and t. It returns a set containing only the elements of s that are not in t.

```
Set set_difference(Set s, Set t)|
   s = s AND set_complement(t)
   return s
```

• Set set_complement(Set s)—: This function takes one parameter of type set: s. It returns the complement set s. This is the set that contains all the elements of the universal set U that are not in s and contains none of the elements that are in s.

```
Set set_complement(Set s)|
    s = NOT s
    return s
```

Insertion Sort

• void insertion_sort(Stats *stats, int *arr, int length): This function takes three parameters: a stats pointer stats, and integer pointer arr, and an integer length. It does not return anything. Its purpose is to sort the elements of arr.

```
void insertion_sort(Stats *stats, int *arr, int length)
  for i, 1 to i < length
    j = i
    temp = move(stats, arr[i])
  while j >= 1 and comp(stats, temp, arr[j-1]) < 0
    arr[j] = move(stats, arr[j-1])
    j = j - 1
  arr[j] = move(stats, temp)</pre>
```

Shell Sort

• void shell_sort(Stats *stats, int *A, int n): This function takes three parameters: a stats pointer stats, and integer pointer A, and an integer n. This function does not return anything. Its purpose is to sort the elements of A.

```
void shell_sort(Stats *stats, int *A, int n)
  for i, 0 to length of gaps array
    for j, gaps[i] to j < length of A array
    k = j</pre>
```

```
temp = move(stats, A[j])
while k >= gap[i] and cmp(stats, temp, A[k - gap[i]]) < 0
    A[k] = move(stats, A[k - gap[i]])
    k = k - gap[i]
A[k] = move(stats, temp)</pre>
```

Heap Sort

• void heap_sort(Stats *stats, int *A, int n): This function takes three parameters: a stats pointer stats, an integer array pointer A, and an integer n. It does not return anything. Its purpose is to sort the elements of array A.

```
void heap_sort(Stats *stats, int *A, int n)
   first = 0
   last = length of A - 1
   build_heap(A, first, last)
   for leaf, last to leaf > first
      swap(stats, A[leaf], A[first])
      fix_heap(A, first, leaf-1)
```

• int max_child(Stats *stats, int *A, int first, int last): This function takes four parameters: a stats pointer stats, an integer array A, an integer first, and an integer last. It returns an integer. It purpose is to determine which element is greater and return it.

```
int max_child(Stats *stats, int *A, int first, int last)
  left = 2 * first + 1
  right = 2 * first + 2
  if right <= last and cmp(stats, A[left], A[right]) < 0 then
      return right
  return left</pre>
```

• void fix_heap(Stats *stats, int *A, int first, int last): This function takes in four parameters: a stats pointer stats, an integer array A, an integer first, and an integer last. It does not return anything. Its purpose is to fix the heap to once again obey the constraints of a max heap after removing the largest element.

```
void fix_heap(Sats *stats, int *A, int first, int last)
  done = false
  parent = first

while ((2 * parent + 1 ) <= last) and done = false
    largest_child = max_child(A, parent, last)
    if cmp(stats, A[parent], A[largest_child]) < 0 then
       swap(stats, A[parent], A[largest_child])
       parent = larges_child
  else
       done = true</pre>
```

• void build_heap(Stats *statsint *A, int first, int last): This function takes in four parameters: a stats pointer stats, a stats pointer stats, an integer array A, an integer first, and an integer last. It does not return anything. Its purpose is to build a max heap.

```
void build_heap(Stats *stats, int *A, int first, int last)
  if last > 0 then
```

```
for parent, (last - 1) / 2 to parent > first - 1
    fix_heap(A, parent, last)
```

Quick Sort

• int partition(Stats *stats, int *A, int lo, int hi): This function takes three parameters, an integer array A, and two integers lo and hi. It returns an integer. Its purpose is to divide the array into two partitions.

```
int partition(Stats *stats, int *A, int lo, int hi)
    i = lo -1
    for j, lo to j < hi
        if cmp(stats, A[j], A[hi]) < 0 then
            i = i + 1
            swap(stats, A[i], A[j])
    i = i + 1
    swap(stats, A[i], A[j])
    return i</pre>
```

• void quick_sorter(Stats *stats, int *A, int lo, int hi): This function takes three parameters, an integer array A, and two integers lo and hi. It does not return anything. Its purpose is to recursively call itself and sort the array A.

```
void quick_sorter(Stats *stats, int *A, int lo, int hi)
   if lo < hi then
      p = partition(stats, A, lo, hi)
      quick_sorter(stats, A, lo, p - 1)
      quick_sorter(stats, A, P + 1, hi)</pre>
```

• void quick_sort(Stats *stats, int *A, int n): This function takes three parameters: a stats pointer stats, an integer array pointer A, and an integer n. It does not return anything. Its purpose is to sort the elements in the array A.

```
void quick_sort(Stats *stats, int *A, int n)
  quick_sorter(stats, A, 0, length of A - 1)
```

Batcher Sort

• void comparator(Stats *stats, int *A, int x, int y): This function takes three parameters: an integer array A, and two integers x and y. It does not return anything. Its purpose is to compare A[x] to A[y] and it it is greater, than swap the values.

```
void comparator(Stats *stats, int *A, int x, int y)
  if cmp(stats, A[y], A[x]) < 0 then
        Swap A[x] and A[y]</pre>
```

• int bit_length(int n): This function takes in an integer n. It returns an integer. Its purpose is to count the number of bits of the number n.

```
int bit_length(int n)
  length = 0
  while n > 0
    n = n >> 1
    length = length + 1
```

```
if 1 = 0 then return 1
return length
```

• void batcher_sort(Stats *stats, int *A, int n): This function takes three parameters: a stats pointer stats, an integer array pointer A, and an integer n. It does not return anything. Its purpose is to sort the array A.

```
void batcher_sort(Stats *stats, int *A, int n)
   if n = 0 the
        return
   t = bit length(n)
   p = 1 \ll (t - 1)
   while p > 0
       q = 1 << (t - 1)
       r = 0
        d = p
          while d > 0
              for i, 0 to i < n - d
                  if (i AND p) = r then
                      comparator(A, i, i + d)
            d = q - p
            q >>= 1
            r = p
       p >>= 1
```

Psuedocode

```
main
    set = 00000000 (0x00)
   heap = 0
    batcher = 1
    shell = 2
    quick = 3
    insert = 4
    elements = 100
    seed = 13371453
    size = 100
   while (opt = getopt) != -1
        switch opt
            case "H":
                print usage message to stdout
                return 0
                break
            case "a":
                set = set_universal()
                break;
            case "h":
                set = set_insert(set, heap)
                break;
            case "b":
                set = set_insert(set, batcher)
                break
```

```
case "s":
            set = set_insert(set, shell)
            break
        case "q":
            set = set_insert(set, quick)
            break
        case "i":
            set = set_insert(set, insert)
            break
        case "n":
            size = optarg
            if size < 0 then
                size = 100
            break
        case "p":
            elements = optarg
            if elements < 0 then
                elements = 100
            break
        case "r":
            seed = optarg
            break
        default:
                print usage message to stderr;
                return 1;
if set == 0x00 then
    print "Select at least one sort to perform" to stderr
    print usage message to stderr
initialize Stats variable stats
mask = 0x3FFFFFFF
srandom(seed)
array = allocated memory of size (size of(first element AND mask) * number of elements)
if set_member(set, insertion) then
   reset(stats)
    init_array(seed, size, array)
    insertion_sort(stats, array, size)
    print_stats(stats, "Insertion Sort", size)
   print_elements
if set_member(set, heap) then
   reset(stats)
    init_array(seed, size, array)
    heapsort(stats, array, size)
    print_stats(stats, "Heap Sort", size)
   print_elements
if set_member(set, shell) then
   reset(stats)
    init_array(seed, size, array)
    shell_sort(stats, array, size)
    print_stats(stats, "Shell Sort", size)
   print_elements
if set_member(set, quick) then
    reset(stats)
```

```
init_array(seed, size, array)
   quick_sort(stats, array, size)
   print_stats(stats, "Quick Sort", size)
   print_elements
if set_member(set, batcher) then
   reset(stats)
   init_array(seed, size, array)
   batcher_sort(stats, array, size)
   print_stats(stats, "Batcher Sort", size)
   print_elements

free(array)
   return 0
```

Error Handling

- Invalid array size: The program checks if the array size is less than 0. If it is, then size is given its default value of 100. If no input is given after, the command line argument -n, an error is printed. If the input does not contain a number, size is set to 0.
- Invalid number of elements: The program checks whether the user's input for the number of elements to print is negative. If it is, the default value of 100 elements will be printed. If the user's input does not contain a number, the program will print zero elements. If the user does not give input, the program will print an error.
- No members in set: If no sorts are chosen through command line arguments, the program the usage message is printed and the program terminates.

Testing

The program will test each sorting algorithm with the randomly generated number. These randomly generated numbers will be bit-masked to fit in 30 bits. The program will be tested with a range of array sizes that show the program is capable of sorting an array up to the size of the available memory. To ensure no memory leaks, the program will be run with valgrind using the commands make debug and valgrind ./sorting. This is shown in Fig. 2. The program was also tested using scan-build make. As seen in Fig.1, no bugs were found.

To confirm the output of the program is correct, the diff command is utilized. To do this, the binary given in the resources repository will be run and the output will be directed into a file. This is done using a command such as ./sorting_ref -a > expect1.txt or ./sorting_ref 2> expect_err1.txt. The same commands are used with this programs executable. The files generated can then be compared using the diff command. This is shown in Fig. 3. The only differences are the executables names in the error message which is as expected.

Results

The program and the sorts within it all work as expected. To compare the sorts to one another, I used a program given by Dr. Veenstra to generate data on the amount of moves and compares each sorting algorithm does with arrays of increasing size and random elements. [2]. Because the cide given by Dr. Veenstra relied on the sorting program, that meant the arrays it sorted would always be randomized. In order to test the sorting algorithms on arrays that were in reverse order, I needed to create an additional program. This program, which I named graph.c, created an array of size n, where n is given by the user as a command argument, with elements in order from largest to smallest. This array was then sorted and remade for each

of the sorting algorithms. The stats of each algorithm were printed out. I then made a program doit2.sh which ran graph.c with increasing n values. I used the same doit.awk file to format the output into a form that could be taken to a spreadsheet. I repeated this process again with a another file that used an already sorted graph with all the sorting algorithms. The graphs produced from this data are shown in Figures 4 - 9. When the arrays consist of random elements, insertion sort performs the worst overall, shell sort is second worst, and the three the other sorts performed similarly with quick sort performing slightly better than the others. Insertion sort performed a similar amount of compares and moves as the other sorts when the array was less than or close to 100 elements. As the array size increased, insertion sort performed marginally more compares and moves than all of the other sorts. This is shown in Fig. 5 and 4. When the arrays elements were in reverse order, quick sort performed equally as bad as insertion sort, which performed more moves and compares than when the elements were randomized. All of the sorts performed very similarly when the array was a size of around 100 or less elements. Batcher sort performed the least amount of moves no matter how many elements, but not by a large margin. It also performed the least amount of compares with a small array size and then performed just slightly more compares than heap sort as the array increased in size. Either batcher or heap sort performed best overall when the array was in reverse order. With random elements, batcher and heapsort still perform just as good, but heap sort does more moves when an array is small than other sorts and quick sort does less moves and compares overall. Fig. 6 and 7 show this. When the array is already sorted, quick sort performs by far the most amount of moves and compares, especially when the array is large. Insertion sort performs the least amount of moves and compares no matter the size of the array. The other three sorts are all very similar for both moves and compares no matter the size of the array. Figures 8 and 9 show the sorts when the array is sorted. Taking all this into consideration, I think using either batcher, heap, or quick sort would be the best option to sort an array. If there is a chance the array is already sorted or is in reverse order, I would not use quick sort.

```
scan-build: Using '/usr/lib/llvm-14/bin/clang' for static analysis
/usr/share/clang/scan-build-14/bin/../libexec/ccc-analyzer -Wall -Wextra -Wstrict-prototypes -Werror -pe
dantic -gdwarf-4 -c batcher.c
/usr/share/clang/scan-build-14/bin/../libexec/ccc-analyzer -Wall -Wextra -Wstrict-prototypes -Werror -pe
dantic -gdwarf-4 -c heap.c
/usr/share/clang/scan-build-14/bin/../libexec/ccc-analyzer -Wall -Wextra -Wstrict-prototypes -Werror -pe
dantic -gdwarf-4 -c insert.c
/usr/share/clang/scan-build-14/bin/../libexec/ccc-analyzer -Wall -Wextra -Wstrict-prototypes -Werror -pe
dantic -gdwarf-4 -c quick.c
/usr/share/clang/scan-build-14/bin/../libexec/ccc-analyzer -Wall -Wextra -Wstrict-prototypes -Werror -pe
dantic -gdwarf-4 -c set.c
/usr/share/clang/scan-build-14/bin/../libexec/ccc-analyzer -Wall -Wextra -Wstrict-prototypes -Werror -pe
dantic -gdwarf-4 -c shell.c
/usr/share/clang/scan-build-14/bin/../libexec/ccc-analyzer -Wall -Wextra -Wstrict-prototypes -Werror -pe
dantic -gdwarf-4 -c sorting.c
/usr/share/clang/scan-build-14/bin/../libexec/ccc-analyzer -Wall -Wextra -Wstrict-prototypes -Werror -pe
dantic -gdwarf-4 -c stats.c
/usr/share/clang/scan-build-14/bin/../libexec/ccc-analyzer -lm -o sorting batcher.o heap.o insert.o quic
k.o set.o shell.o sorting.o stats.o
scan-build: Analysis run complete.
scan-build: Removing directory '/tmp/scan-build-2023-11-05-234739-4014-1' because it contains no reports
scan-build: No bugs found.
savila35@cse13s-vm:~/cse13s/asgn4$
```

Figure 1: Scan-build testing

```
savila35@cse13s-vm:~/cse13s/asgn4$ make debug
clang -Wall -Wextra -Wstrict-prototypes -Werror -pedantic -gdwarf-4 -c batcher.c
clang -Wall -Wextra -Wstrict-prototypes -Werror -pedantic -gdwarf-4 -c heap.c
clang -Wall -Wextra -Wstrict-prototypes -Werror -pedantic -gdwarf-4 -c insert.c
clang -Wall -Wextra -Wstrict-prototypes -Werror -pedantic -gdwarf-4 -c quick.c
clang -Wall -Wextra -Wstrict-prototypes -Werror -pedantic -gdwarf-4 -c set.c
clang -Wall -Wextra -Wstrict-prototypes -Werror -pedantic -gdwarf-4 -c shell.c clang -Wall -Wextra -Wstrict-prototypes -Werror -pedantic -gdwarf-4 -c sorting.c
clang -Wall -Wextra -Wstrict-prototypes -Werror -pedantic -gdwarf-4 -c stats.c
clang -lm -o sorting batcher.o heap.o insert.o quick.o set.o shell.o sorting.o stats.o
[savila35@cse13s-vm:~/cse13s/asgn4$ valgrind ./sorting -a -p 0
==4110== Memcheck, a memory error detector
==4110== Copyright (C) 2002-2017, and GNU GPL'd, by Julian Seward et al.
==4110== Using Valgrind-3.18.1 and LibVEX; rerun with -h for copyright info
==4110== Command: ./sorting -a -p 0
==4110==
Insertion Sort, 100 elements, 2741 moves, 2638 compares
Heap Sort, 100 elements, 1755 moves, 1029 compares
Shell Sort, 100 elements, 3025 moves, 1575 compares
Quick Sort, 100 elements, 1053 moves, 640 compares
Batcher Sort, 100 elements, 1209 moves, 1077 compares
==4110==
==4110== HEAP SUMMARY:
==4110==
              in use at exit: 0 bytes in 0 blocks
==4110==
            total heap usage: 2 allocs, 2 frees, 1,424 bytes allocated
==4110==
==4110== All heap blocks were freed -- no leaks are possible
==4110==
==4110== For lists of detected and suppressed errors, rerun with: -s
==4110== ERROR SUMMARY: 0 errors from 0 contexts (suppressed: 0 from 0)
savila35@cse13s-vm:~/cse13s/asgn4$
```

Figure 2: Valgrind testing

Figure 3: diff testing

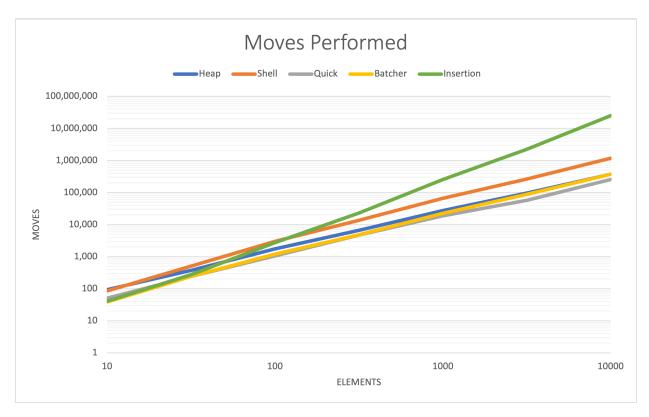


Figure 4: Amount of moves performed

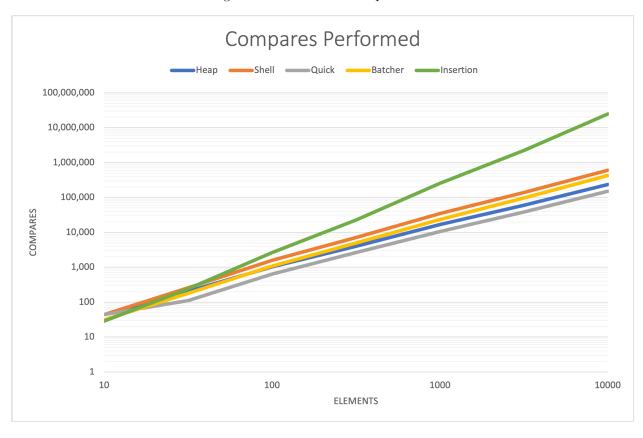


Figure 5: Amount of compares performed

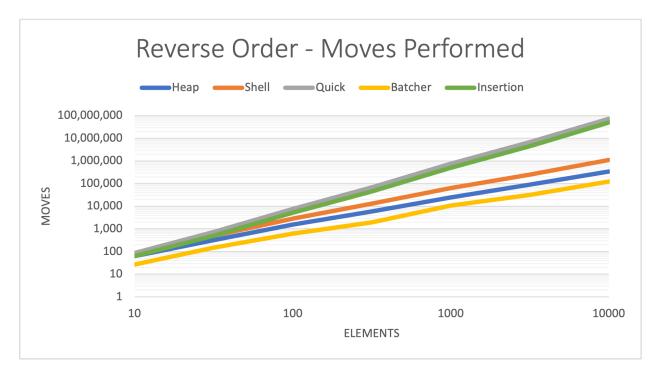


Figure 6: Amount of moves performed when list is reverse sorted

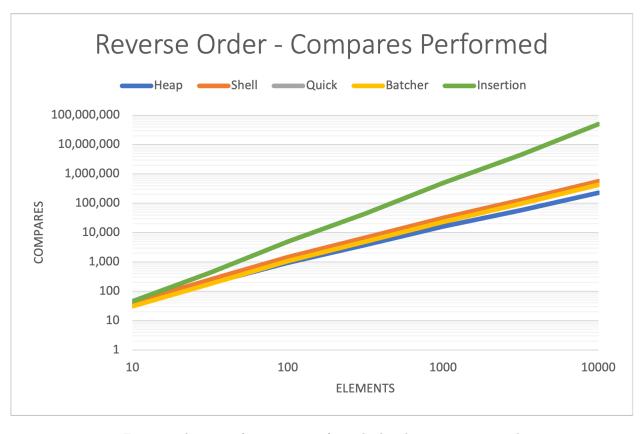


Figure 7: Amount of compares performed when list is reverse sorted

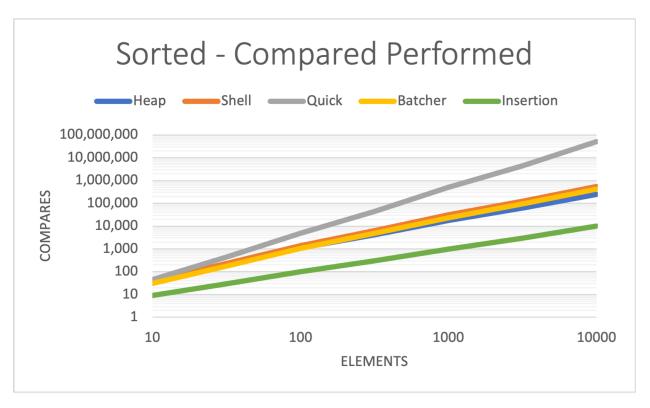


Figure 8: Amount of compares performed when list is already sorted



Figure 9: Amount of compares performed when list is already sorted

References

- [1] Dr. Keery Veenestra and TAs. Assignment 4: Sets and sorting. https://git.ucsc.edu/cse13s/fall-2023-section-01/resources, Fall 2023.
- [2] Dr. Keery Veenestra and TAs. Cse13s canvas page. https://canvas.ucsc.edu/courses/65681, Fall 2023.