Large Scale Data Structures & Organization

Assignment 01: Homework #1

By Saurabh Jawahar Kakade

sk2354@nau.edu

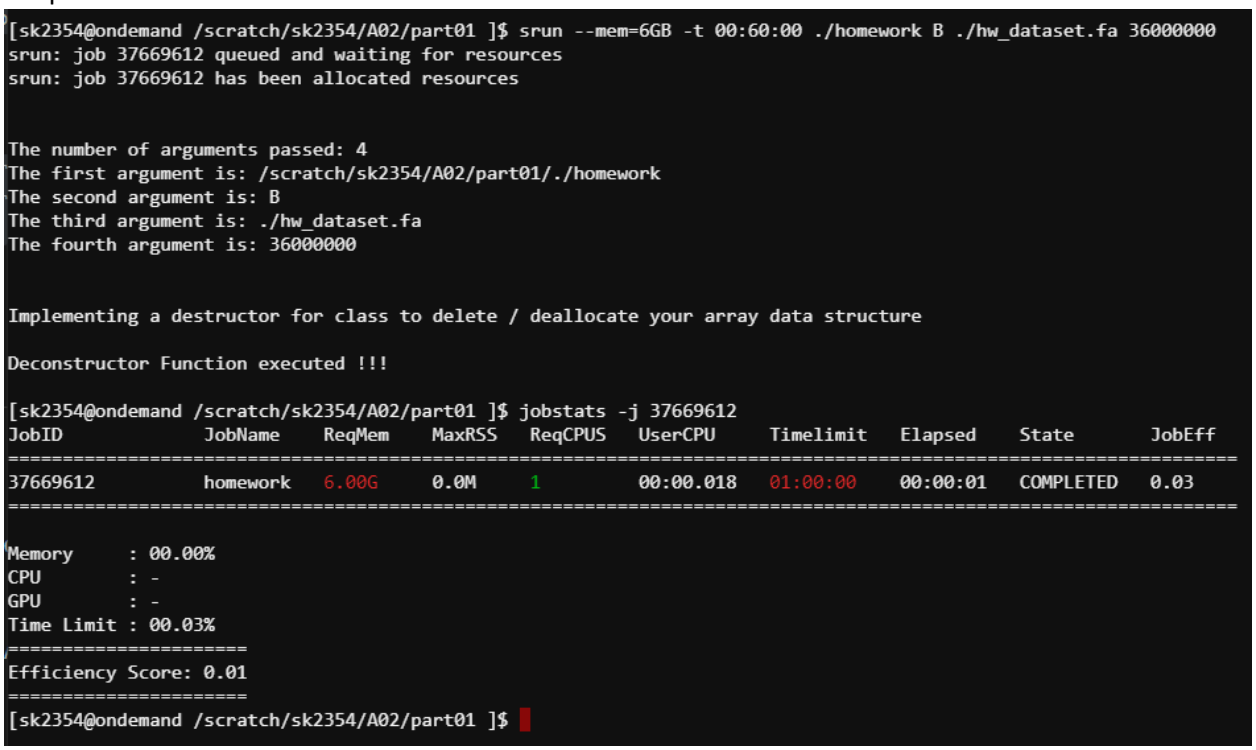
1. Solution:
   Output:

```
[sk2354@ondemand /scratch/sk2354/A02/part01 ]$ srun --mem=6GB -t 00:60:00 ./homework A ./hw_dataset.fa 36000000
srun: job 37669512 queued and waiting for resources
srun: job 37669512 has been allocated resources


The number of arguments passed: 4
The first argument is: /scratch/sk2354/A02/part01/./homework
The second argument is: A
The third argument is: ./hw_dataset.fa
The fourth argument is: 36000000


 Using the first 36 million reads to initialize (fill up) the array data-structure with the entire 36 million reads
[sk2354@ondemand /scratch/sk2354/A02/part01 ]$
[sk2354@ondemand /scratch/sk2354/A02/part01 ]$ jobstats -j 37669512
JobID             JobName    ReqMem    MaxRSS   ReqCPUS   UserCPU     Timelimit   Elapsed    State       JobEff
================================================================================================================
37669512          homework   6.00G     0.0M     1         00:07.008   01:00:00    00:00:13   COMPLETED   0.36
================================================================================================================

Memory     : 00.00%
CPU        : -
GPU        : -
Time Limit : 00.36%
=====================
Efficiency Score: 0.18
=====================
[sk2354@ondemand /scratch/sk2354/A02/part01 ]$
```

- CPU usage: 13 seconds
- RAM: 0.0M

2. Solution:

Output:



Time complexity: O(n)

Explanation: To delete entire list, we need to delete each nodes of the linked list at a time. First create a temp node pointing to second node and delete first head node. Reset temp to head node and repeat this action till we reach NULL. Hence to traverse all nodes to delete them individually, time complexity is O(n).

3. Solution:

   Output:

```
[sk2354@ondemand /scratch/sk2354/A02/part01 ]$ srun --mem=6GB -t 00:60:00 ./homework C ./hw_dataset.fa 36000000
srun: job 37669722 queued and waiting for resources
srun: job 37669722 has been allocated resources


The number of arguments passed: 4
The first argument is: /scratch/sk2354/A02/part01/./homework
The second argument is: C
The third argument is: ./hw_dataset.fa
The fourth argument is: 36000000


Implementation of Copy Constructor:
Copy Constructor Called !!!
[sk2354@ondemand /scratch/sk2354/A02/part01 ]$ jobstats -j 37669722
JobID             JobName    ReqMem    MaxRSS   ReqCPUS   UserCPU     Timelimit   Elapsed    State       JobEff
================================================================================================================
37669722          homework   6.00G     0.0M     1         00:06.992   01:00:00    00:00:13   COMPLETED   0.36
================================================================================================================

Memory     : 00.00%
CPU        : -
GPU        : -
Time Limit : 00.36%
======================
Efficiency Score: 0.18
======================
[sk2354@ondemand /scratch/sk2354/A02/part01 ]$
```

   Time complexity: O(n) because every node of linked list are copied by copy constructor one after the another. For entire list, it becomes O(n)

4. Solution:

   Output:

```
[sk2354@ondemand /scratch/sk2354/A02/part01 ]$ srun --mem=6GB -t 00:60:00 ./homework D ./hw_dataset.fa 36000000
srun: job 37669823 queued and waiting for resources
srun: job 37669823 has been allocated resources


The number of arguments passed: 4
The first argument is: /scratch/sk2354/A02/part01/./homework
The second argument is: D
The third argument is: ./hw_dataset.fa
The fourth argument is: 36000000


Implementing Search function:
Search function for sequence 01CTAGGTACATCCACACACAGCAGCGCATTATGTATTTATTGGATTTATTT

CTAGGTACATCCACACACAGCAGCGCATTATGTATTTATTGGATTTATTT found at index 81

Search function for sequence 02 GCGCGATCAGCTTCGCGCGCACCGCGAGCGCCGATTGCACGAAATGGCGC

Not found

Search function for sequence 03 CGATGATCAGGGGCGTTGCGTAATAGAAACTGCGAAGCCGCTCTATCGCC

Not found

Search function for sequence 04 CGTTGGGAGTGCTTGGTTTAGCGCAAATGAGTTTTCGAGGCTATCAAAAA

CGTTGGGAGTGCTTGGTTTAGCGCAAATGAGTTTTCGAGGCTATCAAAAA found at index 1524

Not found

Search function for sequence 05 ACTGTAGAAGAAAAAAGTGAGGCTGCTCTTTTACAAGAAAAAGTNNNNNN

ACTGTAGAAGAAAAAAGTGAGGCTGCTCTTTTACAAGAAAAAGTNNNNNN found at index 1523

Not found

[sk2354@ondemand /scratch/sk2354/A02/part01 ]$ jobstats -j 37669823
JobID             JobName    ReqMem    MaxRSS   ReqCPUS   UserCPU     Timelimit   Elapsed    State       JobEff
================================================================================================================
37669823          homework   6.00G     0.0M     1         00:09.587   01:00:00    00:00:17   COMPLETED   0.47
================================================================================================================
```

```
[sk2354@ondemand /scratch/sk2354/A02/part01 ]$ jobstats -j 37669823
JobID             JobName    ReqMem    MaxRSS   ReqCPUS   UserCPU     Timelimit   Elapsed    State       JobEff
================================================================================================================
37669823          homework   6.00G     0.0M     1         00:09.587   01:00:00    00:00:17   COMPLETED   0.47
================================================================================================================

Memory     : 00.00%
CPU        : -
GPU        : -
Time Limit : 00.47%
=====================
Efficiency Score: 0.23
=====================
[sk2354@ondemand /scratch/sk2354/A02/part01 ]$
```

Part 02:

a) Output:

Total 50-character fragments: **104545**

```
[sk2354@ondemand /scratch/sk2354/A02/part01 ]$ srun --mem=6GB -t 00:60:00 ./homework E ./test_genome.fasta 36000000
srun: job 37806138 queued and waiting for resources
srun: job 37806138 has been allocated resources


The number of arguments passed: 4
The first argument is: /scratch/sk2354/A02/part01/./homework
The second argument is: E
The third argument is: ./test_genome.fasta
The fourth argument is: 36000000

Part 02 function:

Total 50-character fragments: 104545
srun: error: cn69: task 0: Segmentation fault (core dumped)
[sk2354@ondemand /scratch/sk2354/A02/part01 ]$
```