HOMEWORK-1

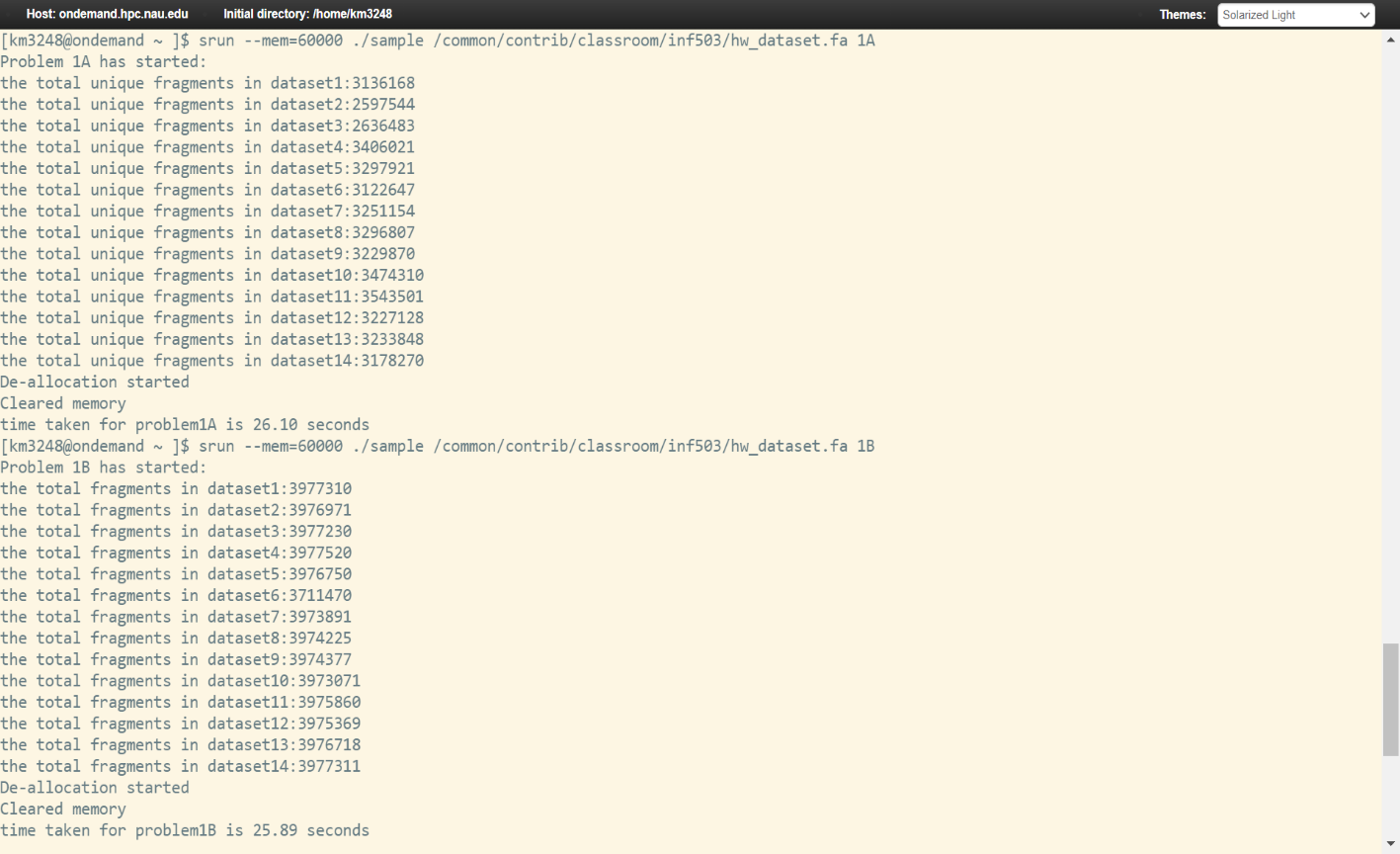
ARRAYS AND CLASSES

BY- Krushith reddy musku,

Km3248@nau.edu

1. How many unique sequence fragments are in each of the 14 datasets?

How many total sequence fragments are in each dataset?



2. Without alphabetically sorting any of the data in the FASTA\_readset object compare the contents of datasets 1 and 2 (i.e. use the fragments in dataset 1 as queries to search in dataset 2). Make sure you continue to consider copy count in your answer.

i. What is the ‘big O’ notation of your search (linear / quadratic / cubic / etc)?

The O(n^2)

ii.How long does it take (in seconds) to search for all fragments of dataset 1 within dataset 2? Please note that depending on the efficiency of your algorithm, this step may take a long time. First estimate the total time using 1,000, 10,000, and 100,000 queries – if total time estimate is greater than 24 CPU hours, provide estimate rather than exact number.

For 1000 queries-> 0.00 seconds

10000 queries-> 0.22 seconds

100000 queries-> 21.59 seconds

For 1Million queries-> 2215.84 seconds

iii.How many sequence fragments in dataset 1 are also in dataset 2?



Alphabetically sort the sequence fragments in each of the FASTA\_readset objects and implement a binary search function to compare the contents of datasets 1 and 2 (i.e. use the fragments in dataset 1 as queries to search in dataset 2).

What is the ‘big O’ notation of your search (linear / quadratic / cubic / etc)?

Since I used binary search the Time complexity is O(logn)

How long (in seconds) does it take to search for 1000 queries? How about 10,000 or 100,000? Does the time increase make sense? Explain the differences (if any) when compared to search times obtained as part of 1B.

For 1000 queries -> 0.005 seconds

10000 queries -> 0.495 seconds

100000 queries-> 66.95 seconds

As you can see there is increase in time that’s because bubble sorting is done before the binary search is implemented. So for bubble sort the time complexity is O(n^2) and for binary search O(logn)

How many sequence fragments in dataset 1 are also in dataset 2?

