**Homework #2**

**Part – A:**

* How long did it take you to search for the first 10K, 100K, and 1M 32-character long fragments of the subject dataset within the query dataset?

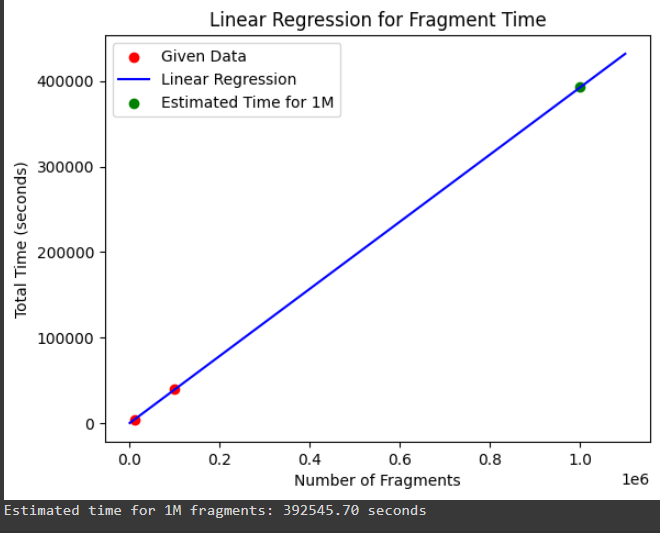
Time taken to search 10k 32-character long fragments: 4057.82 seconds

Time taken to search 100k 32-character long fragments: 39374.9 seconds

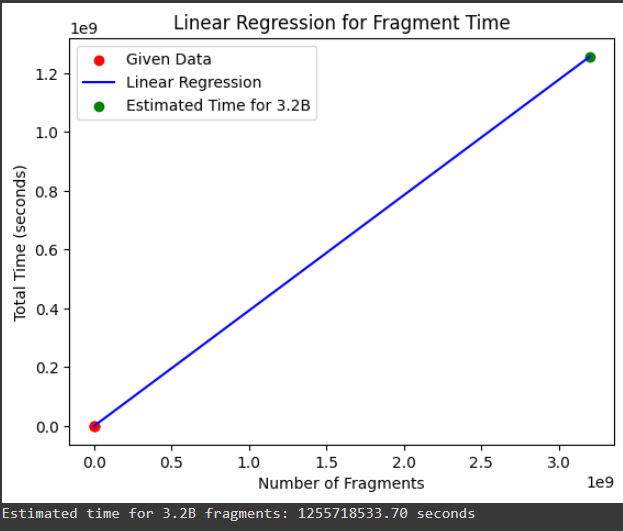
Time taken to search 1M 32-character long fragments:

->It took more than 24 hours, so providing the estimated time to search for 1M 32-character long fragments by taking the time that is taken for 10K,100k fragments as reference.

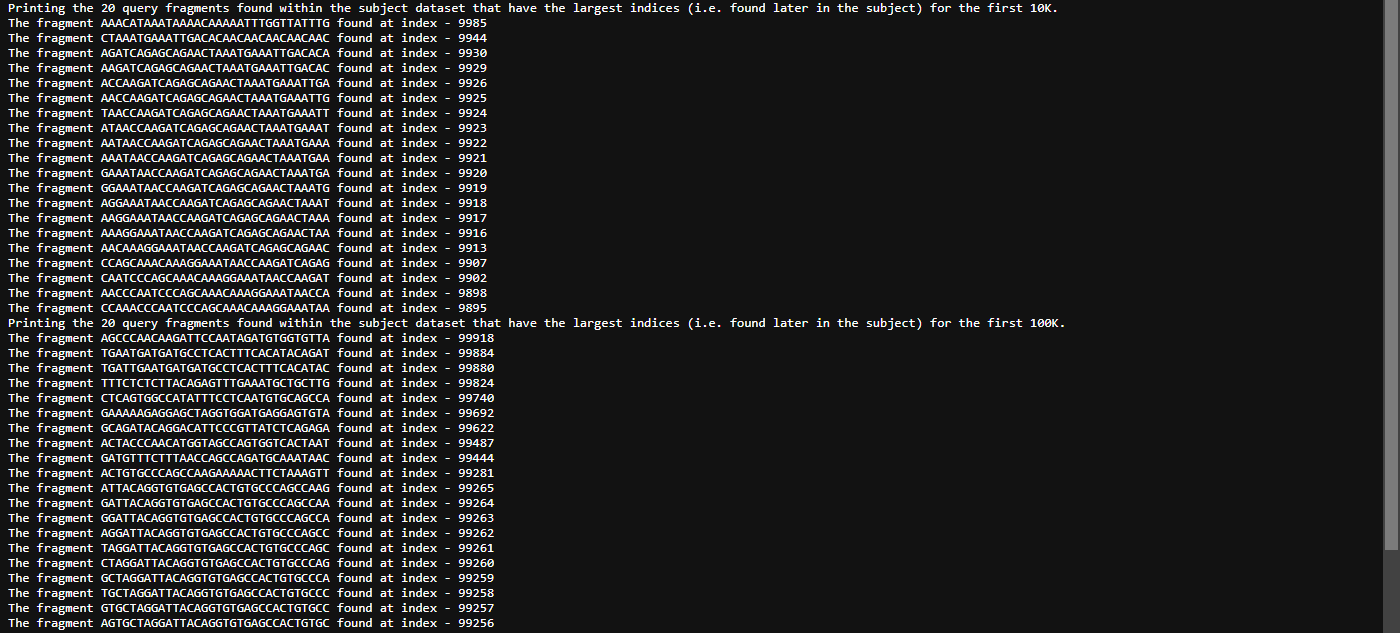
Total estimated time in seconds : 392545.70 seconds

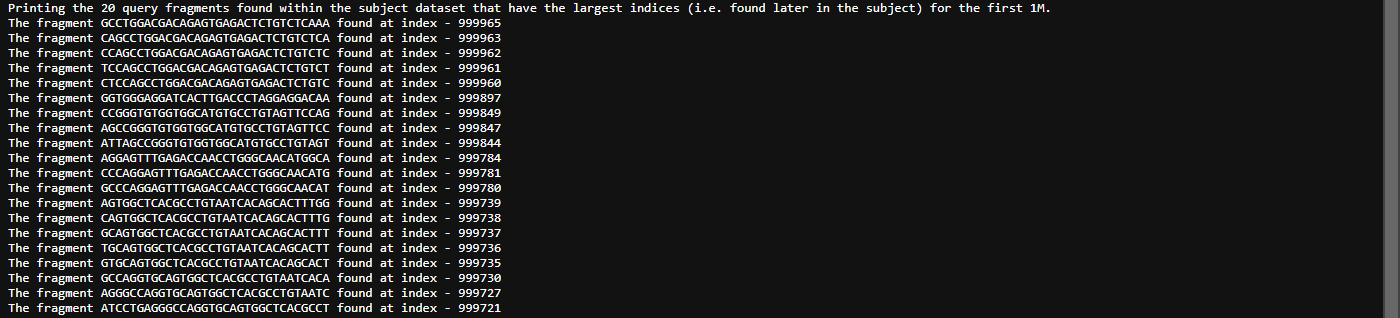


* How long would it take to search for every possible 32-character long fragment of the *subject dataset* within the *query dataset*? Please note that depending on the efficiency of your algorithm, this step may take a long time. If the total time estimate is greater than 24 CPU hours, provide estimate rather than exact number.

Total estimated time in seconds : 1255718533.70 seconds.  
 

* Print the 20 query fragments found within the subject dataset that have the largest indices (i.e. found later in the subject) for the first 10K, 100K, and 1M 32 character long fragments.





**Part-B:**

* How long did it take you to search for the first 10K, 100K, and 1M 32-character long fragments of the subject dataset within the query dataset?

Time taken to search 10k 32-character long fragments: 0.026271 seconds

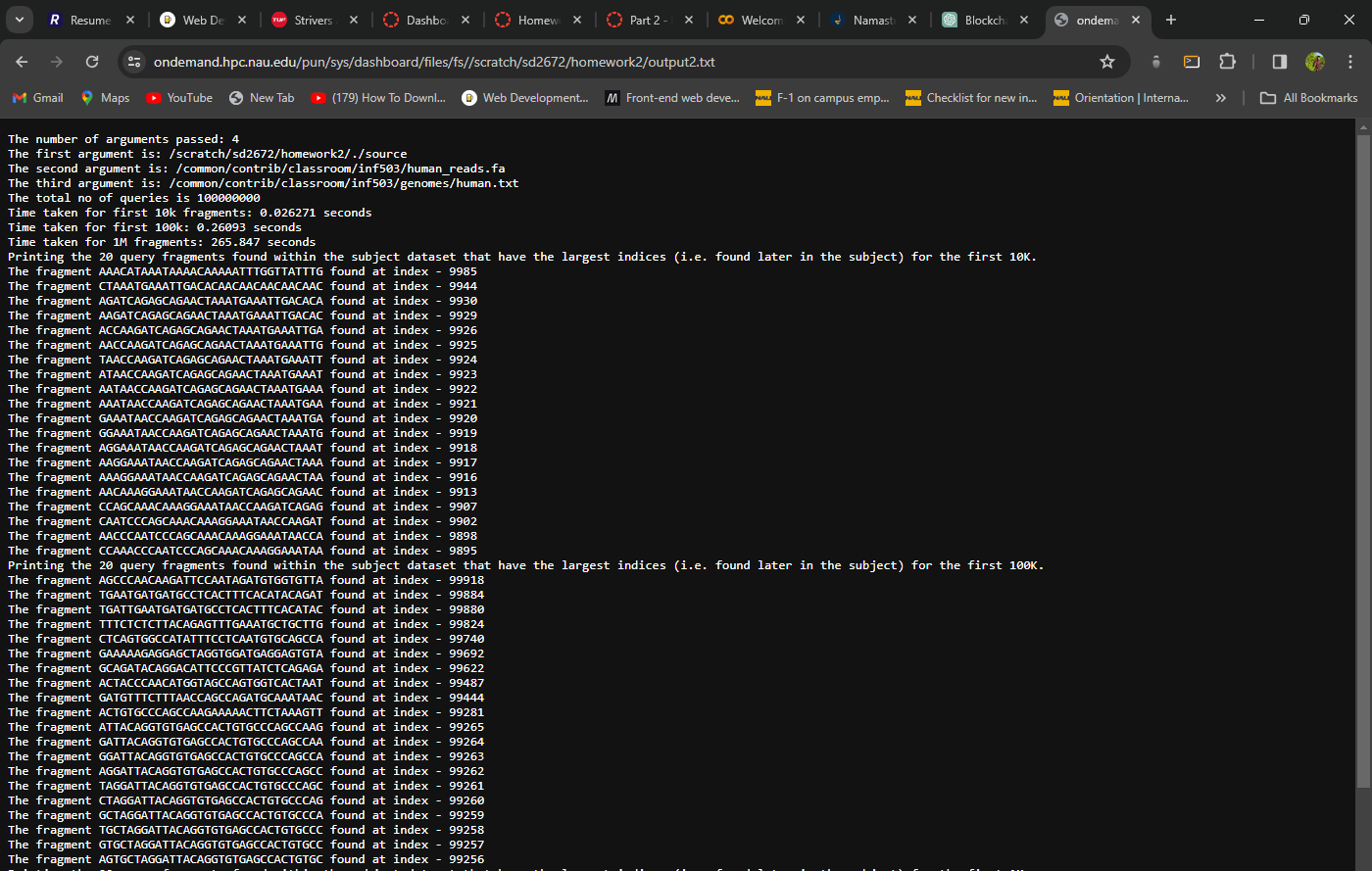
Time taken to search 100k 32-character long fragments: 0.26093 seconds

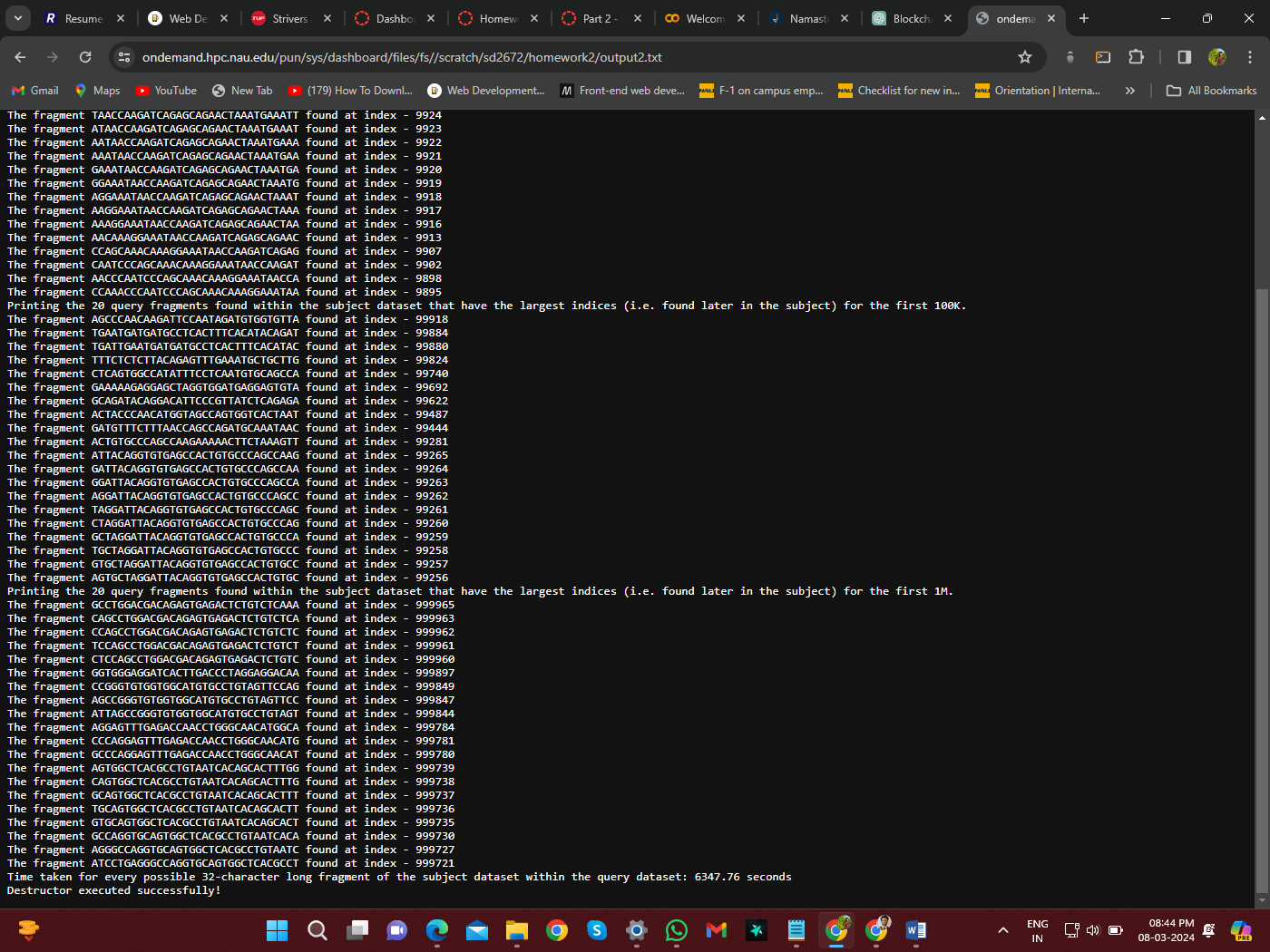
Time taken to search 1M 32-character long fragments: 265.847 seconds

* How long would it take to search for every possible 32-character long fragment of the *subject dataset* within the *query dataset*? Please note that depending on the efficiency of your algorithm, this step may take a long time. If the total time estimate is greater than 24 CPU hours, provide estimate rather than exact number.

6347.76 seconds

* Print the 20 query fragments found within the subject dataset that have the largest indices (i.e. found later in the subject) for the first 10K, 100K, and 1M 32 character long fragments.



****

**Steps of execution:**

* Created total of three files main.cpp, header\_definitions.cpp and header.h
* header.h file contains all the header files that are used in the program
* The main.cpp contains the main function and all the function calls required to get desired output
* Header\_definitions.cpp file contains all the function definitions which are declared in the header file
* Created a make file to run the code
* Uploaded all the above files to a directory on monsoon
* There I have opened terminal and entered the command “make” then source executable file is generated next we need to run the source file with file path and the part of execution.
* The command for execution of part A is

./source /common/contrib/classroom/inf503/human\_reads.fa /common/contrib/classroom/inf503/genomes/human.txt part-A

The command for execution of part B is

./source /common/contrib/classroom/inf503/human\_reads.fa /common/contrib/classroom/inf503/genomes/human.txt part-B