Homework #2

**Name**: Purnabhishek Sripathi

**Email:** [ps747@nau.edu](mailto:ps747@nau.edu)

**Userid:** 6274051

**Problem #1 (of 1)**

1. A default constructor (that zeroes everything out)

A screenshot of a computer

Description automatically generated

1. At least one custom constructor (e.g. one taking a file path or ifstream as input)

A screenshot of a computer

Description automatically generated

1. A function to read the query dataset file.

A screenshot of a computer

Description automatically generated

1. A search function designed to find a sequence fragment within class’s data.

A screenshot of a computer

Description automatically generated

1. A function to sort the fragments of the Queries\_AR object.

A screenshot of a computer code

Description automatically generated

1. A destructor

A screenshot of a computer

Description automatically generated

**A. (30 pts)**

1. 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*?

**Ans:**

**Time taken to search for the first 10K:**

18624 seconds

Approximately it takes 5.1733333 hours.

**Time taken to search for the first 100K:**

181741 seconds

Approximately it takes 50.4836111 hours.

**Time taken to search for the first 1M:** 504.836111 hoursapprox**.**

I have calculated the approximate calculation for 1M is as below:

Time taken to search for the first 1M = (Time taken for the first 100K fragments) \* (1M / 100K)

Time taken to search for first 1M fragments = 50.4836111 hours \* (1,000,000 / 100,000)

Time taken to search for first 1M fragments = 50.4836111 hours \* 10

Time taken to search for first 1M fragments = 504.836111 (**21 days approximately**)

1. **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 is greater than 24 CPU hours, provide an estimate rather than an exact number.**

**Ans:**

The time required for every possible 32-character long fragment within the query data set, it takes approximately **1543378.21 hours**.

To translate it in days it would take **64307.43 days** and if we converted it into years, it would take **176.18 years** to compute.

As I have used Linear search here takes these many years to run the program.

1. **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.**

**Screenshot that shows the last 20 fragments for 10k:**

A screenshot of a computer

Description automatically generated

**Screenshot that shows the last 20 fragments for 10k:**

A screenshot of a computer

Description automatically generated

**For 1M last 20 fragments:**

The program is still running for the 1M 32-character long fragments. Due to this reason, I do not have the last 20 fragments for the 1M character long fragments.

**B. (30 pts)**

**1: 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*?**

**Ans:**

**Time taken to search for the first 10K:** 0.074871 seconds.

**Time taken to search for the first 100K:** 0.625749 seconds.

**Time taken to search for the first 1M:** 4.95256 seconds.

**2: 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 an estimate rather than exact number.**

**Total time taken is:** 13519.3 secondsand it is approximatel**y 3.7552778 hours.**

A screenshot of a computer

Description automatically generated

**3: 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.**

**Ans:**

**Below is the screenshot for the 20 query fragments found within the first 10K.**

A screenshot of a computer

Description automatically generated

**Below is the screenshot for the 20 query fragments found within the first 100K.**

A screenshot of a computer

Description automatically generated

**Below is the screenshot for the 20 query fragments found within the first 1M.**

A screenshot of a computer

Description automatically generated