

Program Structure and Algorithms (INFO 6205)

Fall – 2021 Final Project

Report

Keyur Ashokbhai Barot – 001568664

I. Tasks:

Following tasks were performed in the project:

1. Implement MSD radix sort for a natural language which uses Unicode characters.
2. Sort Simplified Chinese words and compare the results of MSD radix with Timsort, Dual Pivot Quicksort, Huskysort and LSD radix sort.

II. Results:

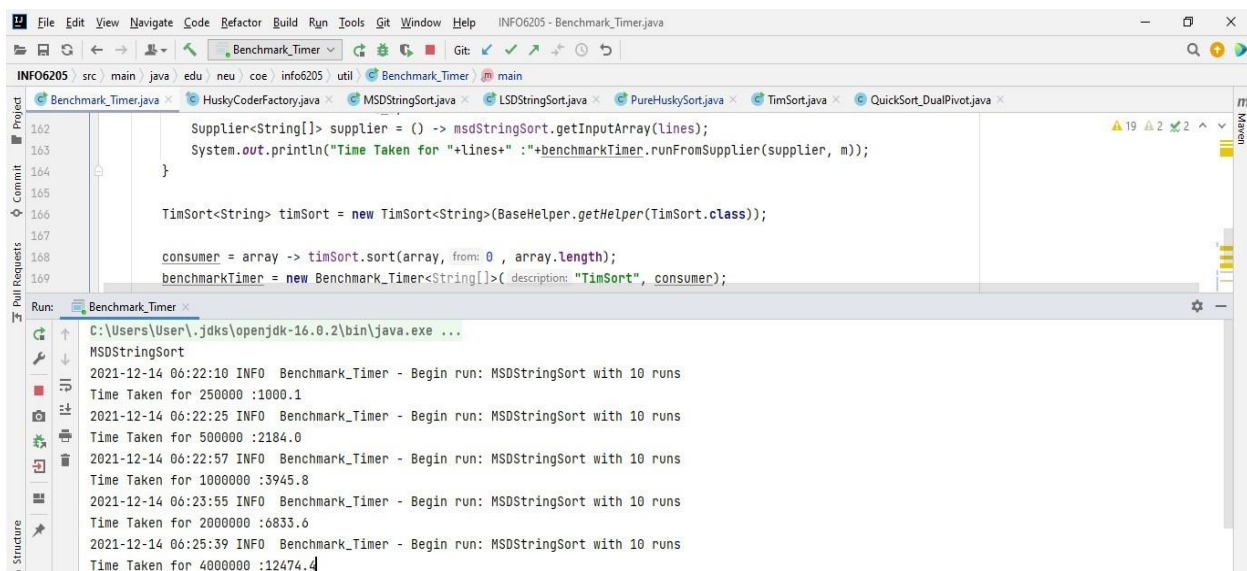
The code for the benchmarking can be found here:

https://github.com/KeyurAshokbhaiBarot/INFO6205/blob/Fall2021/src/main/java/edu/neu/coe/info6205/util/Benchmark_Timer.java

1. MSD Radix Sort:

<https://github.com/KeyurAshokbhaiBarot/INFO6205/blob/Fall2021/src/main/java/edu/neu/coe/info6205/sort/counting/MSDStringSort.java>

The benchmarking results for the MSD radix sort are as follows:



The screenshot shows an IDE window titled 'INFO6205 - Benchmark_Timer.java'. The code in the editor includes a supplier method that calls `msdStringSort.getInputArray(lines)` and prints the time taken. It also creates a `TimSort` instance and sorts an array. The output window at the bottom shows the execution results for `MSDStringSort` with 10 runs, displaying the time taken for each run.

```
Supplier<String[]> supplier = () -> msdStringSort.getInputArray(lines);
System.out.println("Time Taken for "+lines+" :"+benchmarkTimer.runFromSupplier(supplier, m));

TimSort<String> timSort = new TimSort<String>(BaseHelper.getHelper(TimSort.class));

consumer = array -> timSort.sort(array, from 0 , array.length);
benchmarkTimer = new Benchmark_Timer<String[]>(<description> "TimSort", consumer);
```

Run: Benchmark_Timer

```
C:\Users\User\.jdk\openjdk-16.0.2\bin\java.exe ...
MSDStringSort
2021-12-14 06:22:10 INFO Benchmark_Timer - Begin run: MSDStringSort with 10 runs
Time Taken for 250000 :1000.1
2021-12-14 06:22:25 INFO Benchmark_Timer - Begin run: MSDStringSort with 10 runs
Time Taken for 500000 :2184.0
2021-12-14 06:22:57 INFO Benchmark_Timer - Begin run: MSDStringSort with 10 runs
Time Taken for 1000000 :3945.8
2021-12-14 06:23:55 INFO Benchmark_Timer - Begin run: MSDStringSort with 10 runs
Time Taken for 2000000 :6833.6
2021-12-14 06:25:39 INFO Benchmark_Timer - Begin run: MSDStringSort with 10 runs
Time Taken for 4000000 :12474.4
```

Algorithm	Number of words	Time(ms)
MSDStringSort	250000	1000.1
MSDStringSort	500000	2184
MSDStringSort	1000000	3945.8
MSDStringSort	2000000	6833.6
MSDStringSort	4000000	12474.4

2. Timsort:

<https://github.com/KeyurAshokbhaiBarot/INFO6205/blob/Fall2021/src/main/java/edu/neu/coe/info6205/sort/linearithmic/TimSort.java>

The benchmarking results for Timsort are as follows:

The screenshot shows an IDE with the following code in `Benchmark_Timer.java`:

```

162 Supplier<String[]> supplier = () -> msdStringSort.getInputArray(lines);
163 System.out.println("Time Taken for "+lines+" :"+benchmarkTimer.runFromSupplier(supplier, m));
164 }
165
166 TimSort<String> timSort = new TimSort<String>(BaseHelper.getHelper(TimSort.class));
167
168 consumer = array -> timSort.sort(array, from: 0 , array.length);
169 benchmarkTimer = new Benchmark_Timer<String[]>("TimSort", consumer);

```

The Run console shows the following output:

```

Time Taken for 2000000 :6833.6
2021-12-14 06:25:39 INFO Benchmark_Timer - Begin run: MSDStringSort with 10 runs
Time Taken for 4000000 :12474.4
TimSort
2021-12-14 06:28:52 INFO Benchmark_Timer - Begin run: TimSort with 10 runs
Time Taken for 250000 :191.0
2021-12-14 06:28:57 INFO Benchmark_Timer - Begin run: TimSort with 10 runs
Time Taken for 500000 :388.7
2021-12-14 06:29:07 INFO Benchmark_Timer - Begin run: TimSort with 10 runs
Time Taken for 1000000 :794.8
2021-12-14 06:29:28 INFO Benchmark_Timer - Begin run: TimSort with 10 runs
Time Taken for 2000000 :1604.6
2021-12-14 06:30:09 INFO Benchmark_Timer - Begin run: TimSort with 10 runs
Time Taken for 4000000 :3413.8

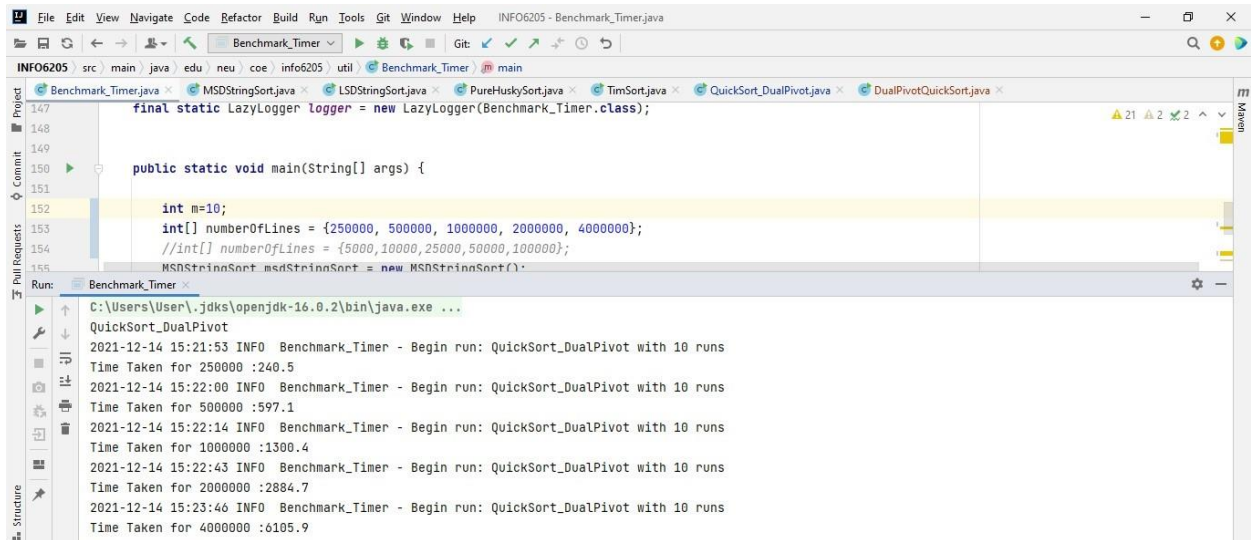
```

Algorithm	Number of words	Time(ms)
TimSort	250000	191
TimSort	500000	388.7
TimSort	1000000	794.8
TimSort	2000000	1604.6
TimSort	4000000	3413.8

3. Dual-pivot Quicksort:

<https://github.com/KeyurAshokbhaiBarot/INFO6205/blob/Fall2021/src/main/java/edu/neu/coe/info6205/sort/DualPivotQuickSort.java>

The benchmarking results for Dual-pivot Quicksort are as follows:



```
final static LazyLogger logger = new LazyLogger(Benchmark_Timer.class);

public static void main(String[] args) {

    int m=10;
    int[] numberofLines = {250000, 500000, 1000000, 2000000, 4000000};
    //int[] numberofLines = {5000,10000,25000,50000,100000};
    MSDStringSort msdSteinnSort = new MSDStringSort();

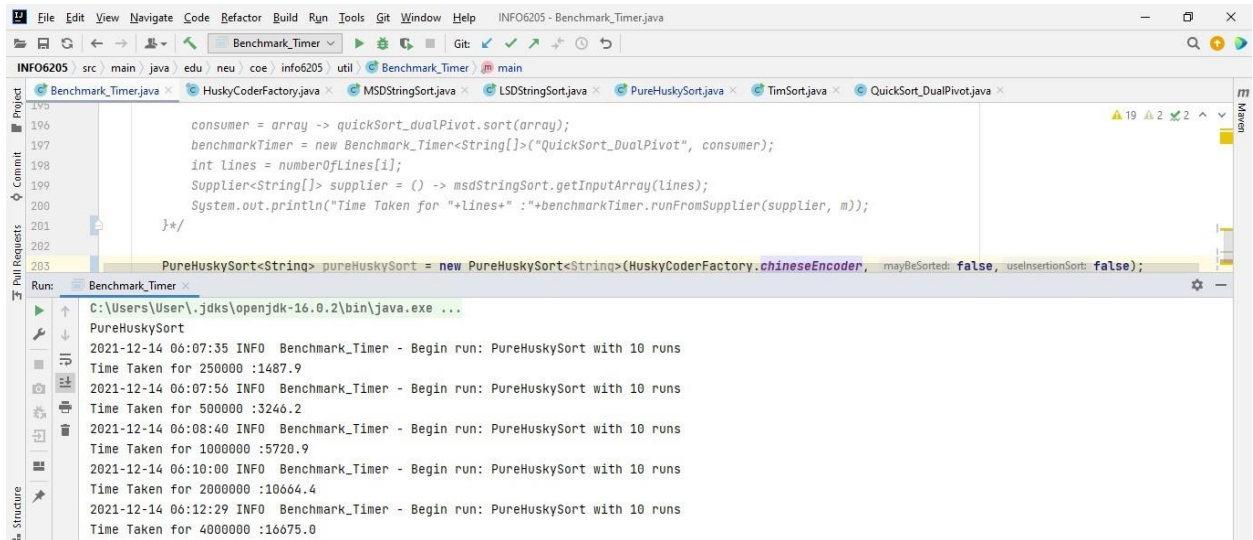
    C:\Users\User\jdk\openjdk-16.0.2\bin\java.exe ...
    QuickSort_DualPivot
    2021-12-14 15:21:53 INFO Benchmark_Timer - Begin run: QuickSort_DualPivot with 10 runs
    Time Taken for 250000 :240.5
    2021-12-14 15:22:00 INFO Benchmark_Timer - Begin run: QuickSort_DualPivot with 10 runs
    Time Taken for 500000 :597.1
    2021-12-14 15:22:14 INFO Benchmark_Timer - Begin run: QuickSort_DualPivot with 10 runs
    Time Taken for 1000000 :1300.4
    2021-12-14 15:22:43 INFO Benchmark_Timer - Begin run: QuickSort_DualPivot with 10 runs
    Time Taken for 2000000 :2884.7
    2021-12-14 15:23:46 INFO Benchmark_Timer - Begin run: QuickSort_DualPivot with 10 runs
    Time Taken for 4000000 :6105.9
```

Algorithm	Number of words	Time(ms)
QuickSort_DualPivot	250000	176.6
QuickSort_DualPivot	500000	446
QuickSort_DualPivot	1000000	1069.7
QuickSort_DualPivot	2000000	2302.9
QuickSort_DualPivot	4000000	4971.9

4. Huskysort:

<https://github.com/KeyurAshokbhaiBarot/INFO6205/blob/Fall2021/src/main/java/edu/neu/coe/info6205/sort/huskySort/PureHuskySort.java>

The benchmarking results for Huskysort are as follows:



```
consumer = array -> quickSort_dualPivot.sort(array);
benchmarkTimer = new Benchmark_Timer<String[]>("QuickSort_DualPivot", consumer);
int lines = numberOfLines[i];
Supplier<String[]> supplier = () -> msdStringSort.getInputArray(lines);
System.out.println("Time Taken for "+lines+" :"+benchmarkTimer.runFromSupplier(supplier, m));
}*/

PureHuskySort<String> pureHuskySort = new PureHuskySort<String>(HuskyCoderFactory.chineseEncoder, mayBeSorted: false, useInsertionSort: false);

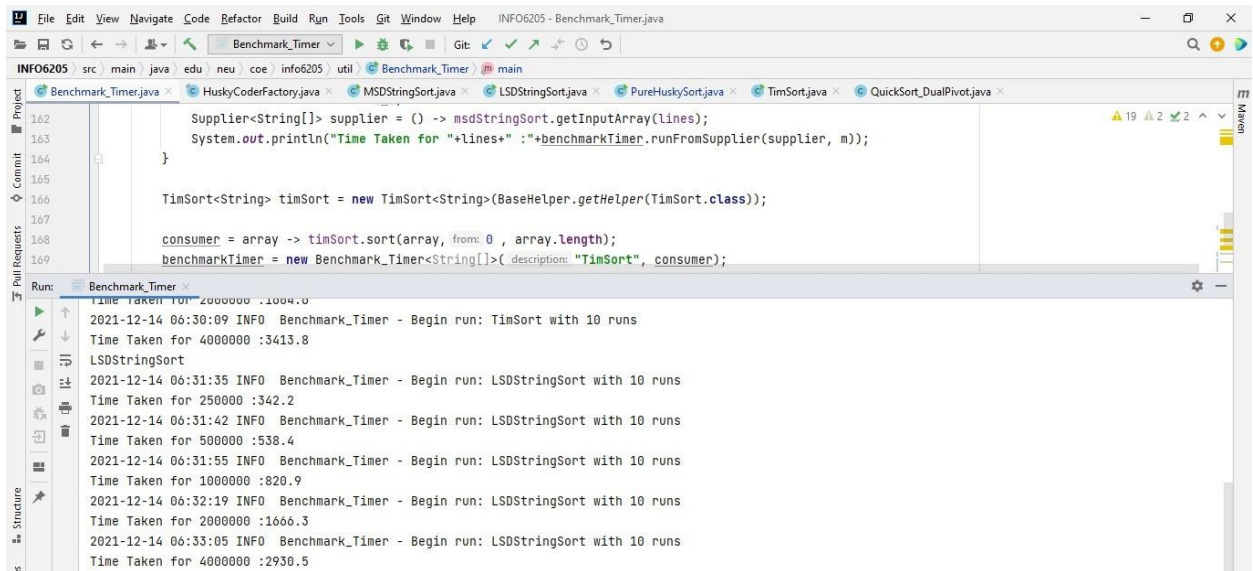
C:\Users\User\jdk\openjdk-16.0.2\bin\java.exe ...
PureHuskySort
2021-12-14 06:07:35 INFO Benchmark_Timer - Begin run: PureHuskySort with 10 runs
Time Taken for 250000 :1487.9
2021-12-14 06:07:56 INFO Benchmark_Timer - Begin run: PureHuskySort with 10 runs
Time Taken for 500000 :3246.2
2021-12-14 06:08:40 INFO Benchmark_Timer - Begin run: PureHuskySort with 10 runs
Time Taken for 1000000 :5720.9
2021-12-14 06:10:00 INFO Benchmark_Timer - Begin run: PureHuskySort with 10 runs
Time Taken for 2000000 :10664.4
2021-12-14 06:12:29 INFO Benchmark_Timer - Begin run: PureHuskySort with 10 runs
Time Taken for 4000000 :16675.0
```

Algorithm	Number of words	Time(ms)
PureHuskySort	250000	623.2
PureHuskySort	500000	1285
PureHuskySort	1000000	2589.8
PureHuskySort	2000000	5223.2
PureHuskySort	4000000	10506.6

5. LSD radix sort:

<https://github.com/KeyurAshokbhaiBarot/INFO6205/blob/Fall2021/src/main/java/edu/neu/coe/info6205/sort/counting/LSDStringSort.java>

The benchmarking results of LSD radix sort are as follows:



```
File Edit View Navigate Code Refactor Build Run Tools Git Window Help INFO6205 - Benchmark_Timer.java
INFO6205 src main java edu neu coe info6205 util Benchmark_Timer main
Benchmark_Timer.java HuskyCoderFactory.java MSDStringSort.java LSDStringSort.java PureHuskySort.java TimSort.java QuickSort_DualPivot.java
162 Supplier<String[]> supplier = () -> msdStringSort.getInputArray(lines);
163 System.out.println("Time Taken for "+lines+" :"+benchmarkTimer.runFromSupplier(supplier, m));
164 }
165
166 TimSort<String> timSort = new TimSort<String>(BaseHelper.getHelper(TimSort.class));
167
168 consumer = array -> timSort.sort(array, from: 0 , array.length);
169 benchmarkTimer = new Benchmark_Timer<String[]>(description: "TimSort", consumer);

Run: Benchmark_Timer
Time Taken for 2000000 :1004.0
2021-12-14 06:30:09 INFO Benchmark_Timer - Begin run: TimSort with 10 runs
Time Taken for 4000000 :3413.8
LSDStringSort
2021-12-14 06:31:35 INFO Benchmark_Timer - Begin run: LSDStringSort with 10 runs
Time Taken for 250000 :342.2
2021-12-14 06:31:42 INFO Benchmark_Timer - Begin run: LSDStringSort with 10 runs
Time Taken for 500000 :538.4
2021-12-14 06:31:55 INFO Benchmark_Timer - Begin run: LSDStringSort with 10 runs
Time Taken for 1000000 :820.9
2021-12-14 06:32:19 INFO Benchmark_Timer - Begin run: LSDStringSort with 10 runs
Time Taken for 2000000 :1666.3
2021-12-14 06:33:05 INFO Benchmark_Timer - Begin run: LSDStringSort with 10 runs
Time Taken for 4000000 :2930.5
```

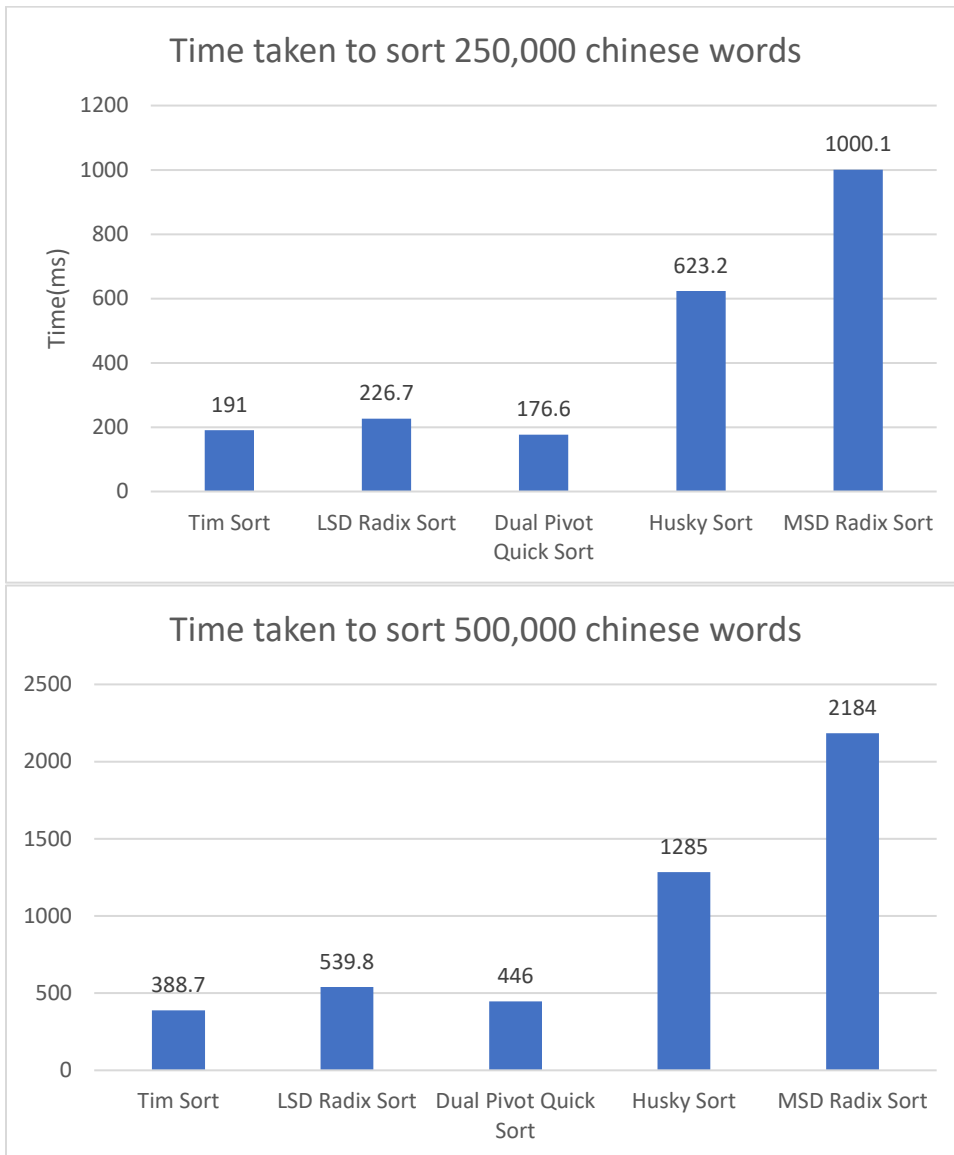
Algorithm	Number of words	Time(ms)
LSDStringSort	250000	226.7
LSDStringSort	500000	539.8
LSDStringSort	1000000	848.5
LSDStringSort	2000000	1503.7
LSDStringSort	4000000	2977.3

III. Conclusion:

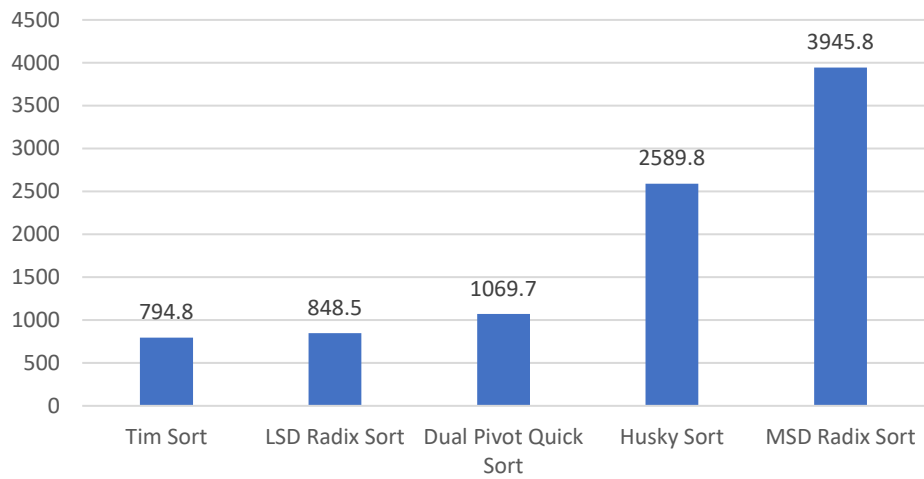
After implementing all the five given algorithms we can say that Timsort takes the least time to sort, followed by LSD radix sort, which is followed by Dual-Pivot Quicksort, Huskysort and MSD radix sort respectively.

Evidence for the same can be seen in the following time comparisons of each algorithm.

IV. Evidence:



Time taken to sort 1,000,000 chinese words



Time taken to sort 2,000,000 chinese words

