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:

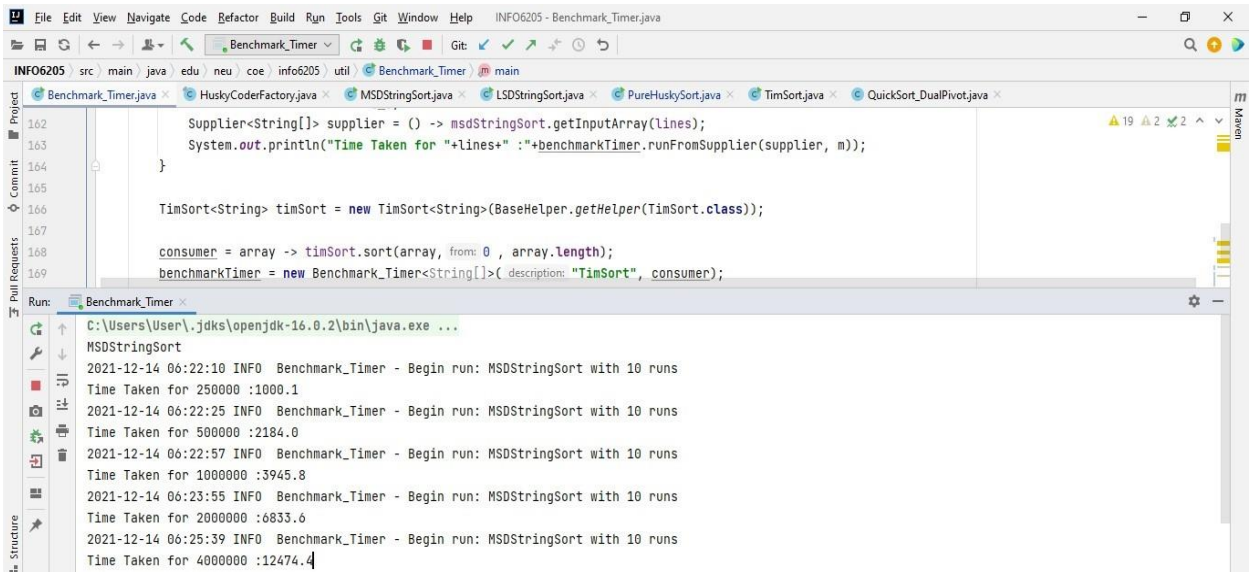
The code for the benchmarking can be found here:

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:

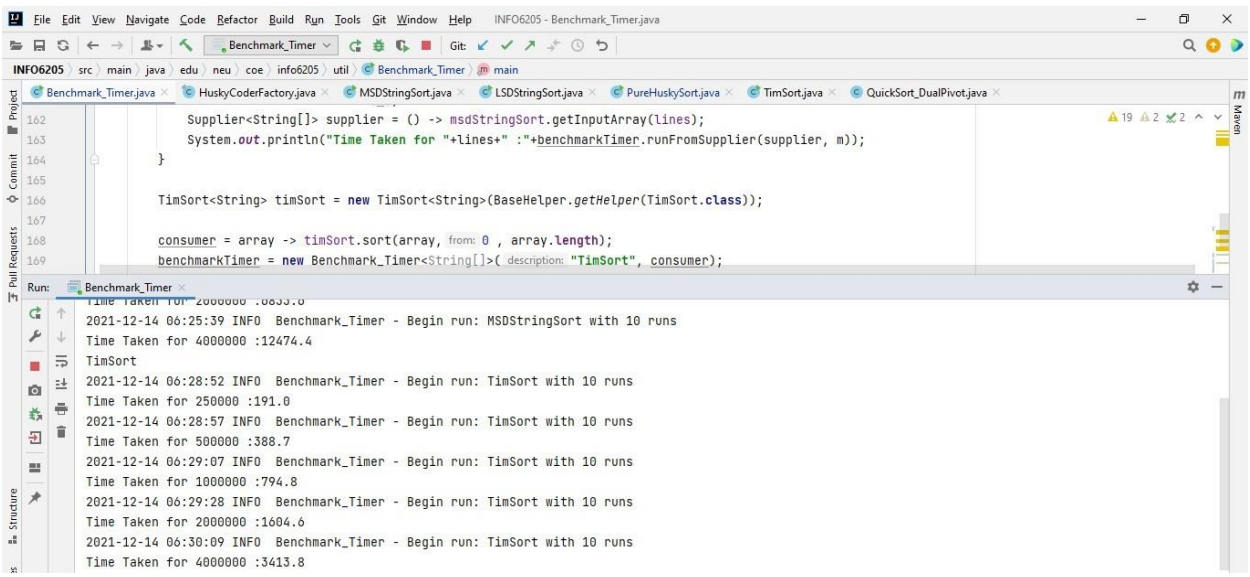


Algorithm	Number of words	Time(ms)
MSDStringSort	250000	1000.1
MSDStringSort	500000	2184
MSDStringSort	1000000	3945.8
MSDStringSort	2000000	6833.6
MSDStringSort	4000000	12474.4

2. Timsort:

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

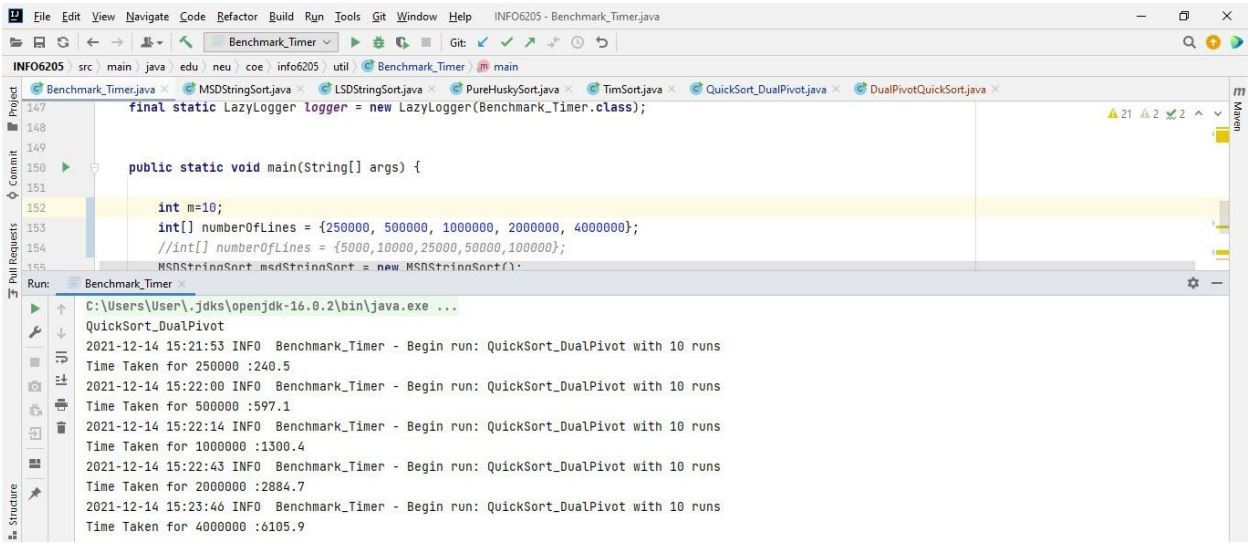
The benchmarking results for Timsort are as follows:



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

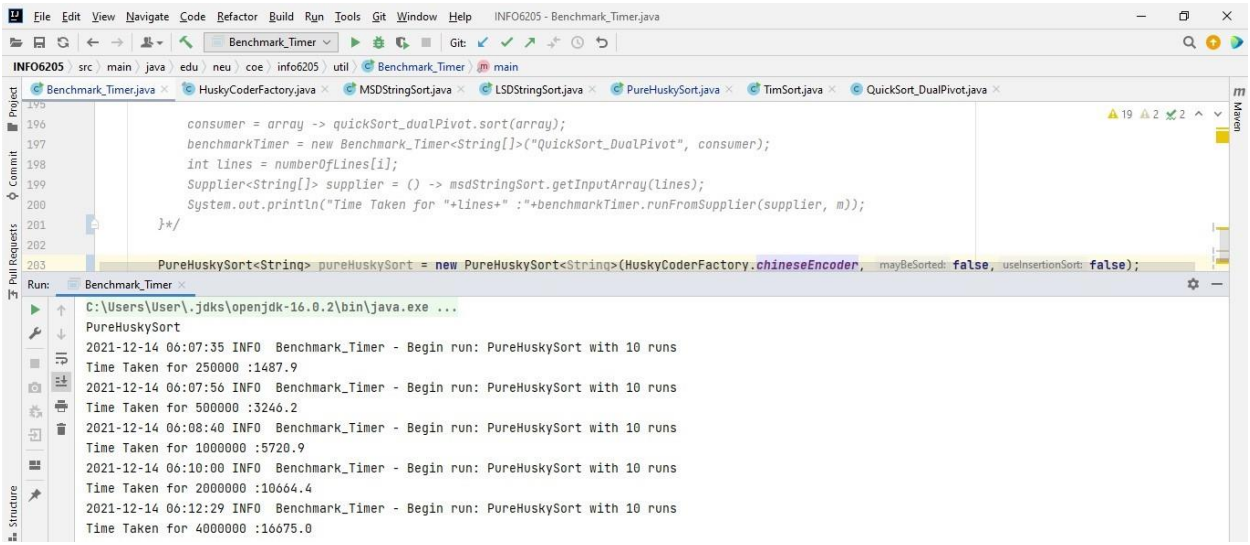
The benchmarking results for Dual-pivot Quicksort are as follows:



Algorithm	Number of words	Time(ms)
QuickSort_DualPivot	250000	240.5
QuickSort_DualPivot	500000	597.1
QuickSort_DualPivot	1000000	1300.4
QuickSort_DualPivot	2000000	2884.7
QuickSort_DualPivot	4000000	6105.9

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:

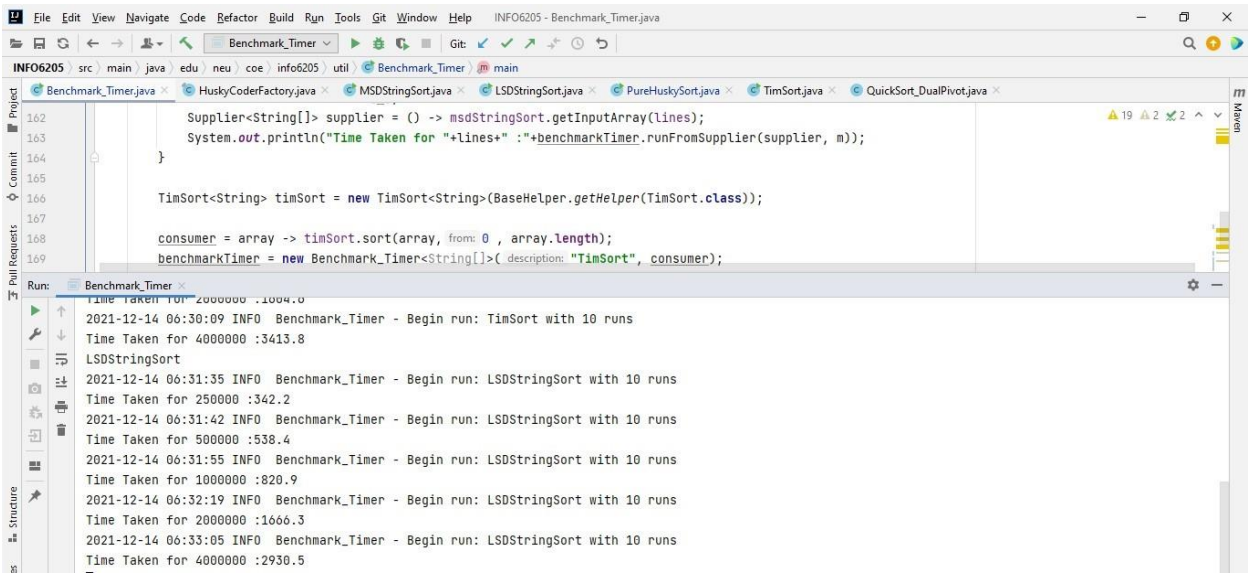


Algorithm	Number of words	Time(ms)
PureHuskySort	250000	1487.9
PureHuskySort	500000	3246.2
PureHuskySort	1000000	5720.9
PureHuskySort	2000000	10664.4
PureHuskySort	4000000	16675

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:



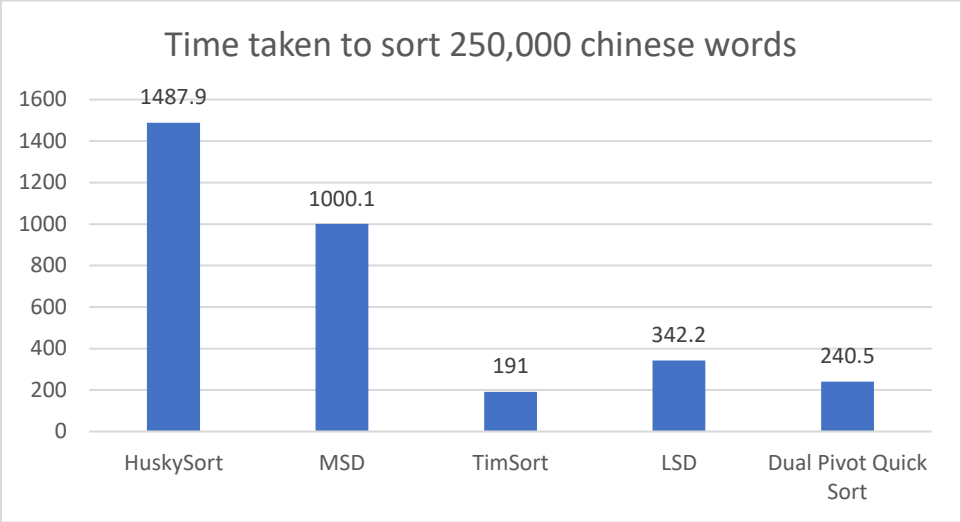
Algorithm	Number of words	Time(ms)
LSDStringSort	250000	342.2
LSDStringSort	500000	538.4
LSDStringSort	1000000	820.9
LSDStringSort	2000000	1666.3
LSDStringSort	4000000	2930.5

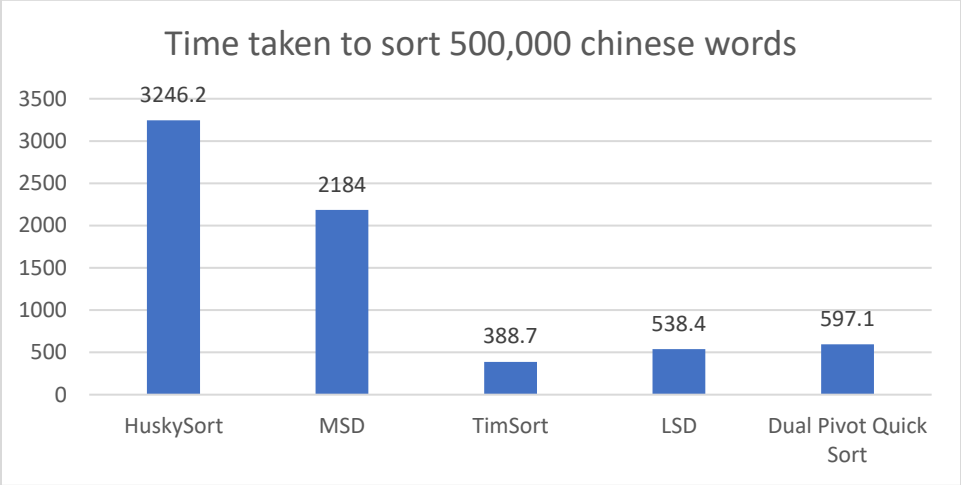
III. Conclusion:

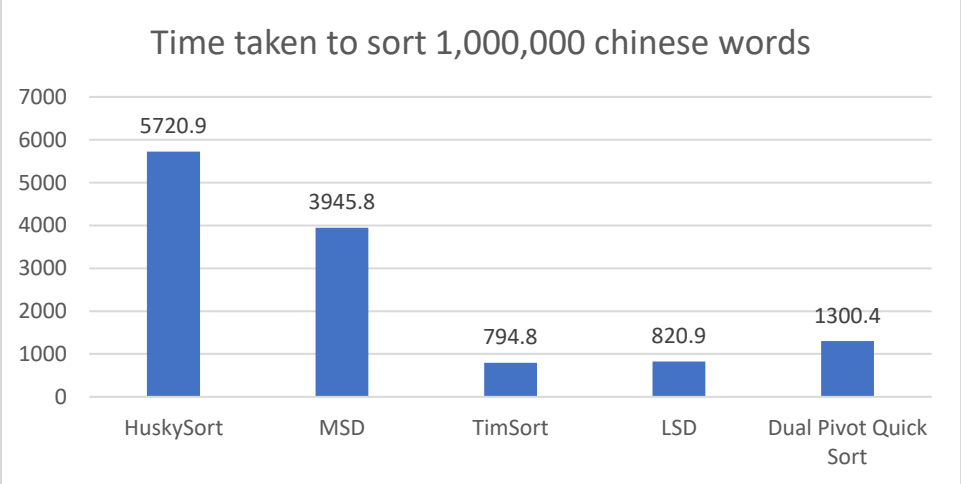
After implementing all the five given algorithms we can say that Timsort takes the least time to sort, followed by LSD radix sort, which is followed by Dual-Pivot Quicksort, MSD radix sort and Huskysort respectively.

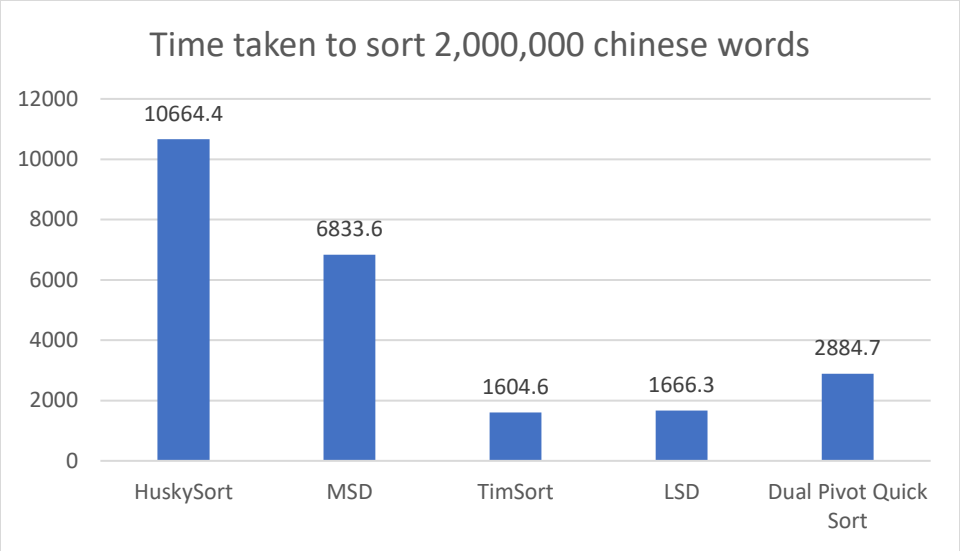
Evidence for the same can be seen in the following time comparisons of each algorithm.

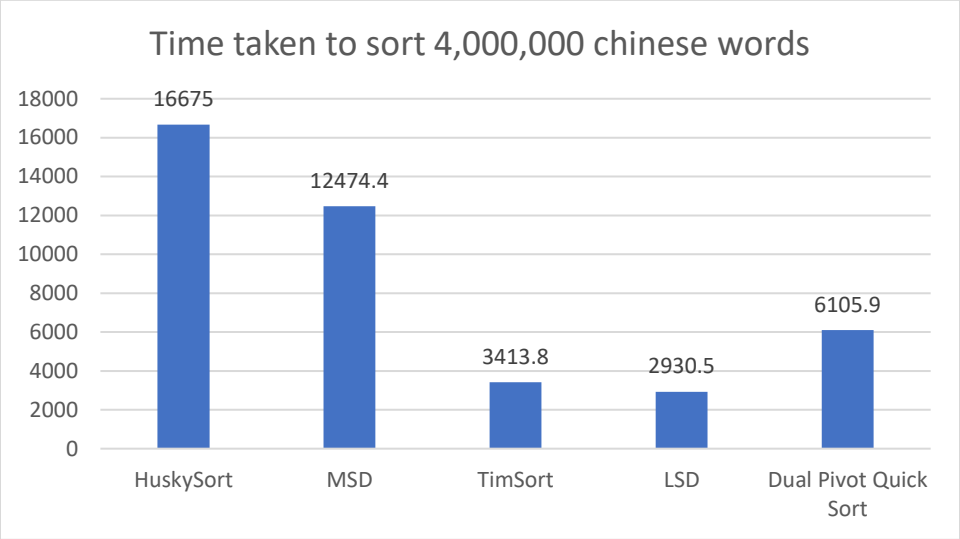
IV. Evidence:


Time taken to sort 250,000 chinese words
1600
1400
1200
1000
800
600
400
200
0
1487.9
1000.1
191
342.2
240.5
HuskySort
MSD
TimSort
LSD
Dual Pivot Quick Sort


Time taken to sort 500,000 chinese words
3500
3000
2500
2000
1500
1000
500
0
3246.2
2184
388.7
538.4
597.1
HuskySort
MSD
TimSort
LSD
Dual Pivot Quick Sort


Time taken to sort 1,000,000 chinese words
7000
6000
5000
4000
3000
2000
1000
0
5720.9
3945.8
794.8
820.9
1300.4
HuskySort
MSD
TimSort
LSD
Dual Pivot Quick Sort


Time taken to sort 2,000,000 chinese words
12000
10000
8000
6000
4000
2000
0
10664.4
6833.6
1604.6
1666.3
2884.7
HuskySort
MSD
TimSort
LSD
Dual Pivot Quick Sort


Time taken to sort 4,000,000 chinese words
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
16675
12474.4
3413.8
2930.5
6105.9
HuskySort
MSD
TimSort
LSD
Dual Pivot Quick Sort