

**Ram Gopal Varma
Venkata Sai Akhil
Mohit Sanghai**

**INFO 6205 Program Structures and Algorithms
Final Project**

Task:

A Report which summarizes your overall findings and recommendations—and shows, graphically, your main conclusions

Observation:

This Report aimed to evaluate the performance of various sorting algorithms used for natural language. Through implementations of Husky sort, Tim sort, Dual-pivot quicksort, MSD radix sort, and three-way radix quicksort, we sorted data types of any type including strings. To conclude the above sorting implementations, we benchmarked them with 1 million Chinese words according to pinyin. We also benchmarked them with other natural languages such as Telugu and English. Several unit tests were added to cover each implementation of the algorithm. Husky sort has been observed to outperform all other algorithms when sorting a million words. We had gone through a few relevant technical papers on the subject of sorting.

Conclusion:

We benchmarked the above sorting algorithms for randomized input of 1 million Chinese words. We sorted them as per the pinyin order. The results are as follows -

Sort Algorithm	Time (in sec)
Husky sort	0.7
Tim sort	5
Dual Pivot Quicksort	7
LSD	21
MSD	71
Three-way Radix Sort	11

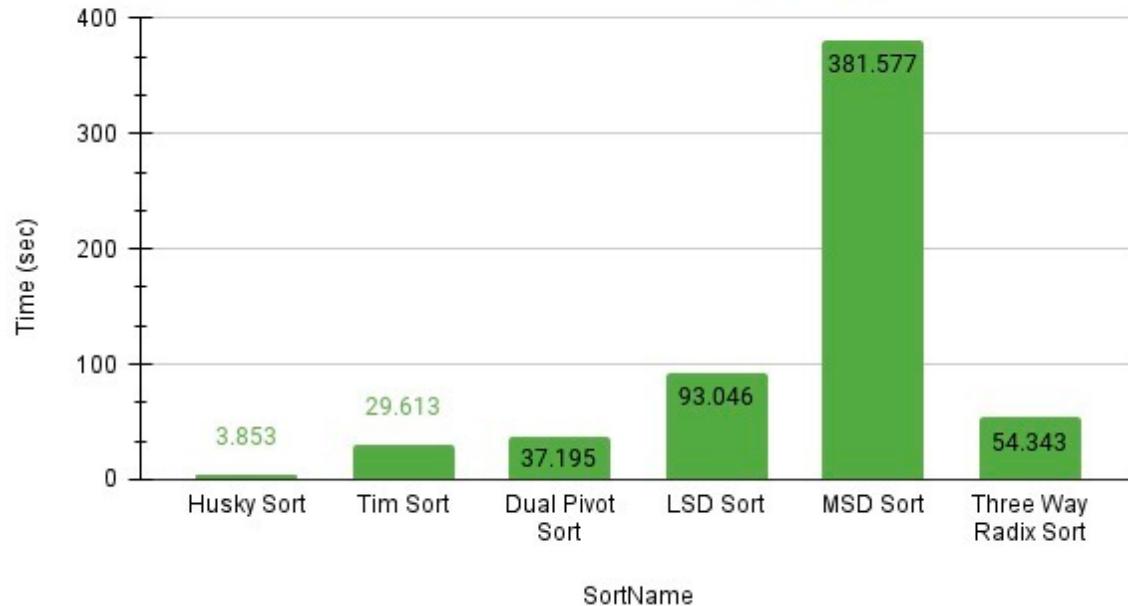
It shows that Husky sort is the best sorting algorithm to sort strings.

Recommendation:

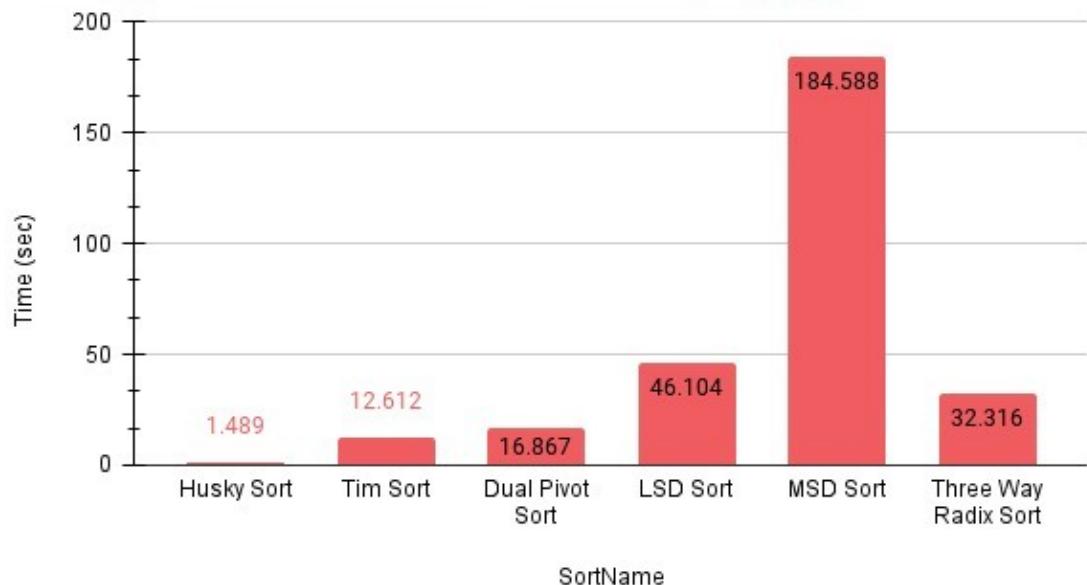
If the array size is huge ,randomised and the type of data is string consider using husky sort. If the data is partially sorted consider using Tim sort

Evidence:

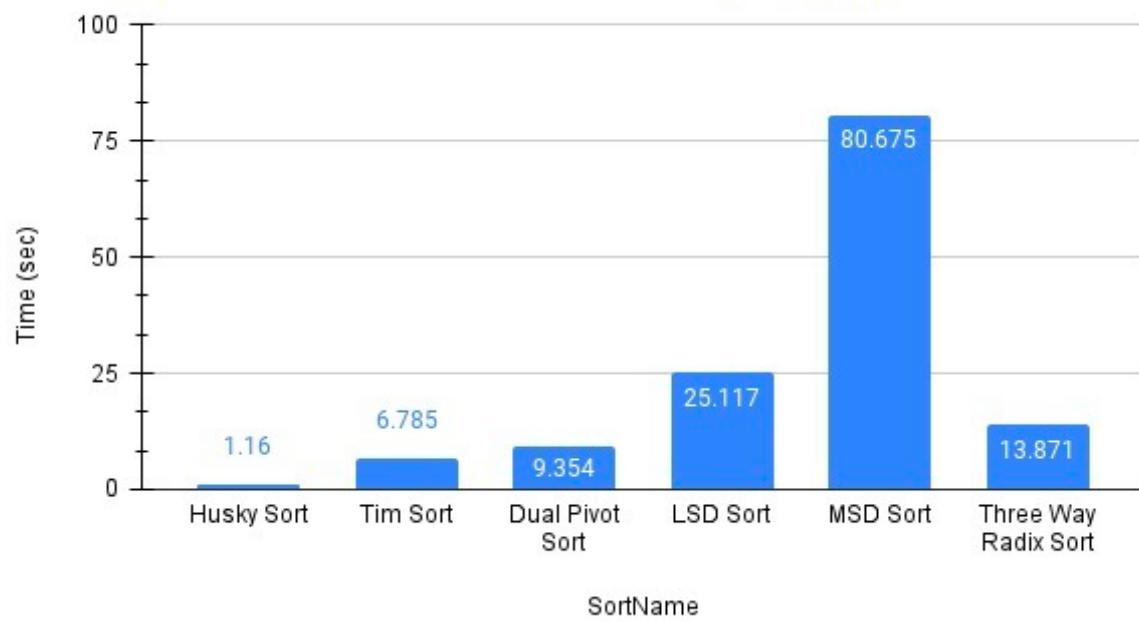
Time (sec) vs SortName for Chinese Words of length 3999992



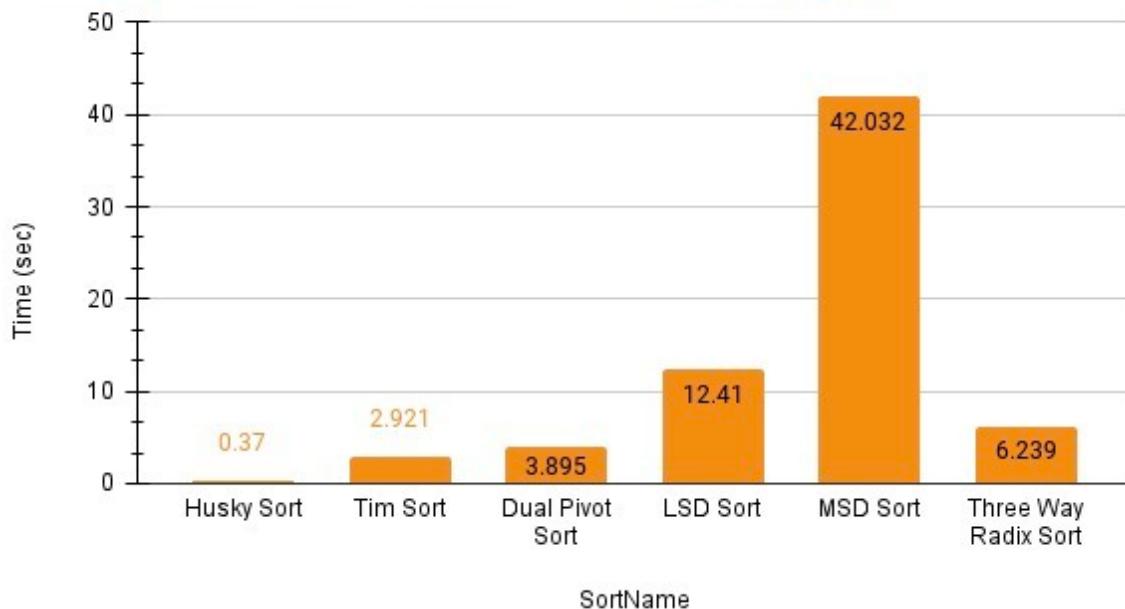
Time (sec) vs SortName for chinese words of length 1999996



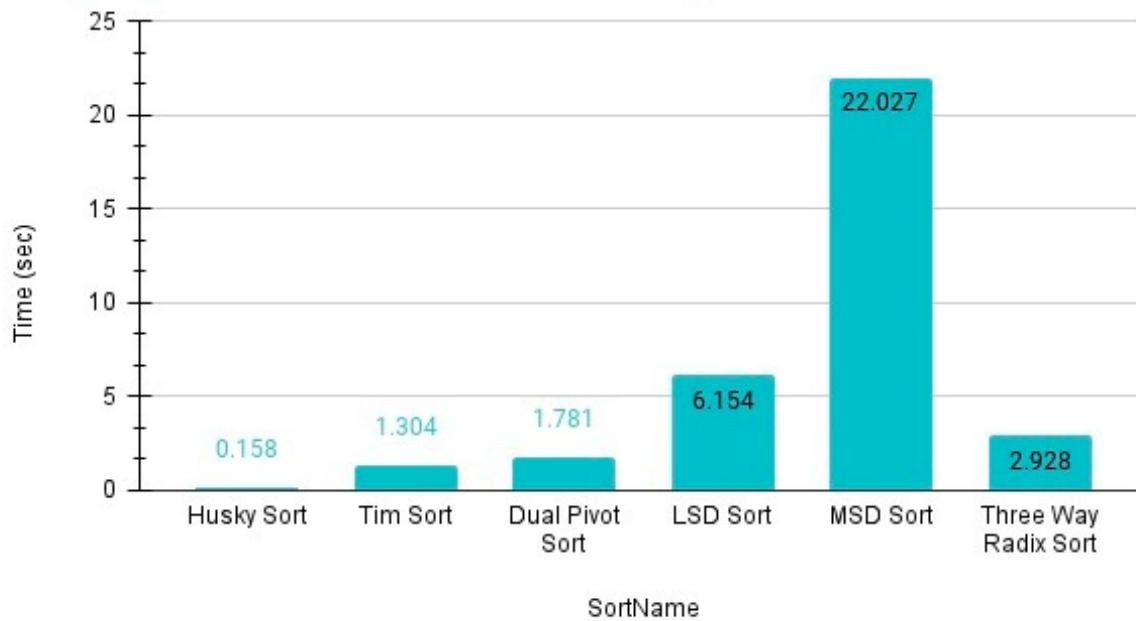
Time (sec) vs SortName for chinese words of length 999998



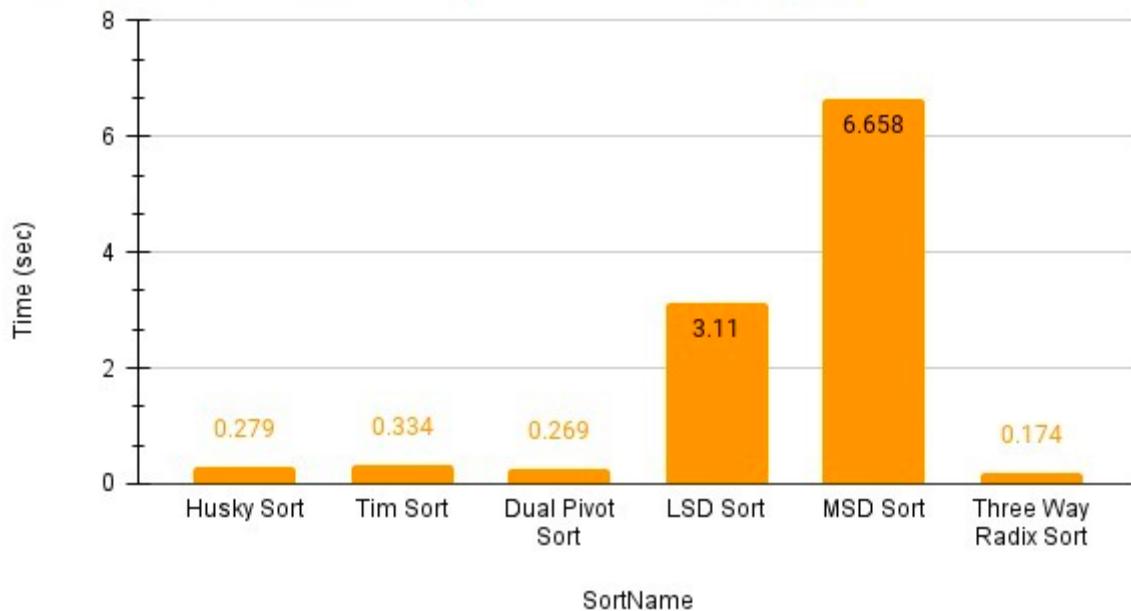
Time (sec) vs SortName for chinese words of length 499999



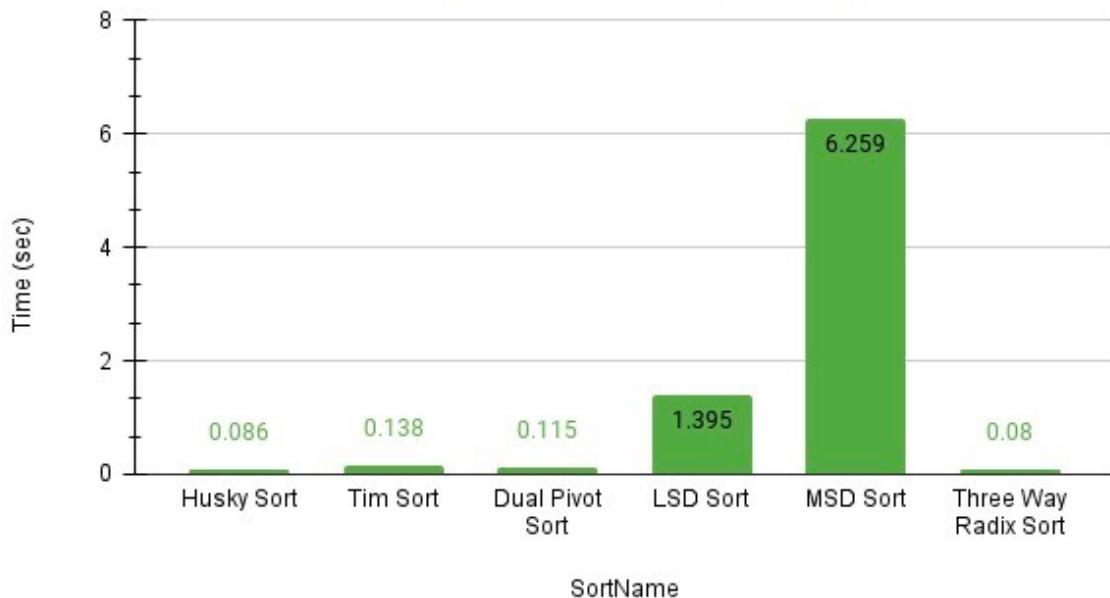
Time (sec) vs SortName for chinese words of length 249999



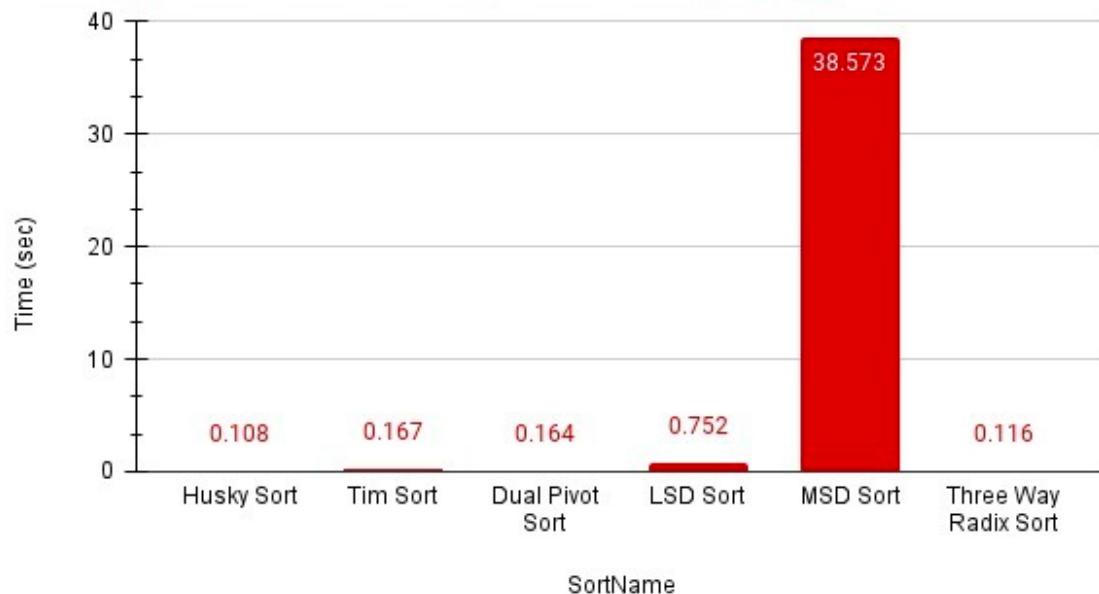
Time (sec) vs SortName for english words of length 442310



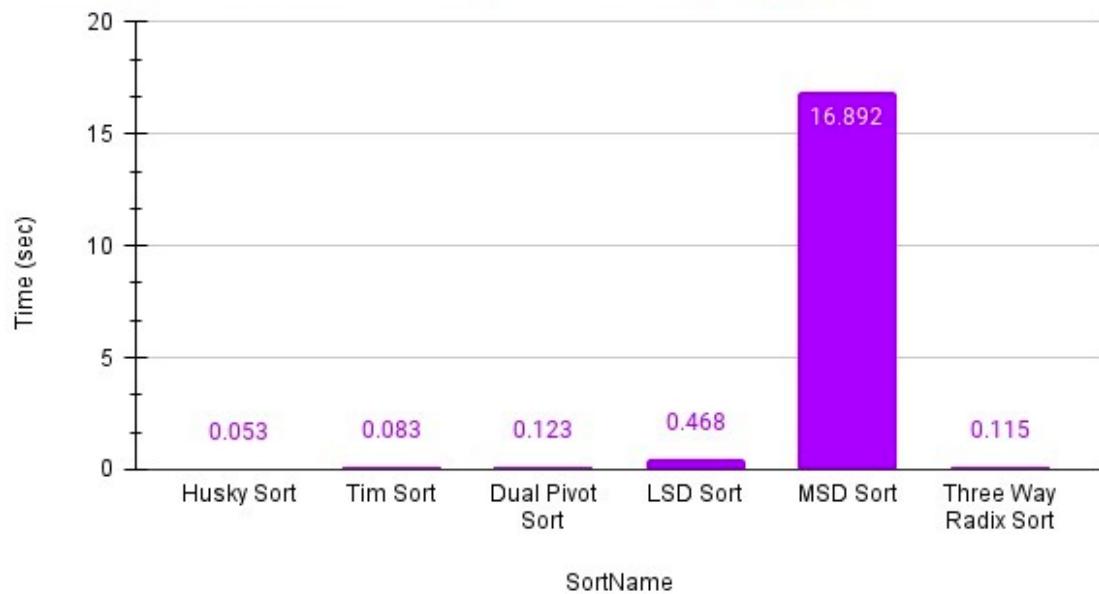
Time (sec) vs SortName for english words of length 221155



Time (sec) vs SortName for Telugu words of length 308568



Time (sec) vs SortName for Telugu words of length 154284



```
INFO6205-FinalProject | src | main | java | edu | neu | coe | info6205 | benchmark | Benchmark_Timer | runFromSupplier
Dec 03, 2021 8:12:38 PM edu.neu.coe.info6205.benchmark.Timer repeat
INFO: repeat: with 10 runs
Husky Sort Average Sort Time = 711.4 ms
Dec 03, 2021 8:12:37 PM edu.neu.coe.info6205.benchmark.Benchmark_Timer runFromSupplier
INFO: Begin run: Tim Sort Test with 10 runs
Dec 03, 2021 8:12:37 PM edu.neu.coe.info6205.benchmark.Timer repeat
INFO: repeat: with 2 runs
Dec 03, 2021 8:12:48 PM edu.neu.coe.info6205.benchmark.Timer repeat
INFO: repeat: with 10 runs
Tie Sort Average Sort Time = 5513.9 ms
Dec 03, 2021 8:13:43 PM edu.neu.coe.info6205.benchmark.Benchmark_Timer runFromSupplier
INFO: Begin run: Dual Pivot Quick Sort Test with 10 runs
Dec 03, 2021 8:13:43 PM edu.neu.coe.info6205.benchmark.Timer repeat
INFO: repeat: with 2 runs
Dec 03, 2021 8:13:57 PM edu.neu.coe.info6205.benchmark.Timer repeat
INFO: repeat: with 10 runs
Dual Pivot Quick Sort Average Sort Time = 7877.8 ms
Dec 03, 2021 8:15:00 PM edu.neu.coe.info6205.benchmark.Benchmark_Timer runFromSupplier
INFO: Begin run: LSD Sort Test with 10 runs
Dec 03, 2021 8:15:00 PM edu.neu.coe.info6205.benchmark.Timer repeat
INFO: repeat: with 2 runs
Dec 03, 2021 8:15:51 PM edu.neu.coe.info6205.benchmark.Timer repeat
INFO: repeat: with 10 runs
LSD String Sort Average Sort Time = 21330.4 ms
Dec 03, 2021 8:19:24 PM edu.neu.coe.info6205.benchmark.Benchmark_Timer runFromSupplier
INFO: Begin run: MSD Sort Test with 10 runs
Dec 03, 2021 8:19:24 PM edu.neu.coe.info6205.benchmark.Timer repeat
INFO: repeat: with 2 runs
Dec 03, 2021 8:21:45 PM edu.neu.coe.info6205.benchmark.Timer repeat
INFO: repeat: with 10 runs
MSD String Sort Average Sort Time = 70547.5 ms
Dec 03, 2021 8:33:48 PM edu.neu.coe.info6205.benchmark.Benchmark_Timer runFromSupplier
INFO: Begin run: Three Way Radix Sort with 10 runs
Dec 03, 2021 8:33:48 PM edu.neu.coe.info6205.benchmark.Timer repeat
INFO: repeat: with 2 runs
Dec 03, 2021 8:34:02 PM edu.neu.coe.info6205.benchmark.Timer repeat
INFO: repeat: with 10 runs
Three way radix Sort Average Sort Time = 11416.0 ms
Process finished with exit code 0
```

[* 2: Git 3: 7000 4: Run 5: Services 6: Messages 7: SonarLint 8: Terminal]

Build completed successfully in 3 s 76 ms (28 minutes ago)

Event Log

Testcases:

IntelliJ IDEA 2021.3.1

Fall2021TeamProject - TimSortTest.java

```
INFO6205-FinalProject src test java edu neu coe info6205 sort TimSortTest
  Project Structure
  Screenshot
  Run: TimSortTest
  Tests passed: 10
```

TimSortTest.java

```
public class TimSortTest {
    @Test
    public void sortInteger0() {
        int n = 20;
        Integer[] input = new Integer[]{4, 3, 5, 6, 1};
        Integer[] expected = Arrays.copyOf(input, input.length);
        Arrays.sort(expected);
        new TimSort<Integer>().sort(input);
        assertEquals(expected, input);
    }

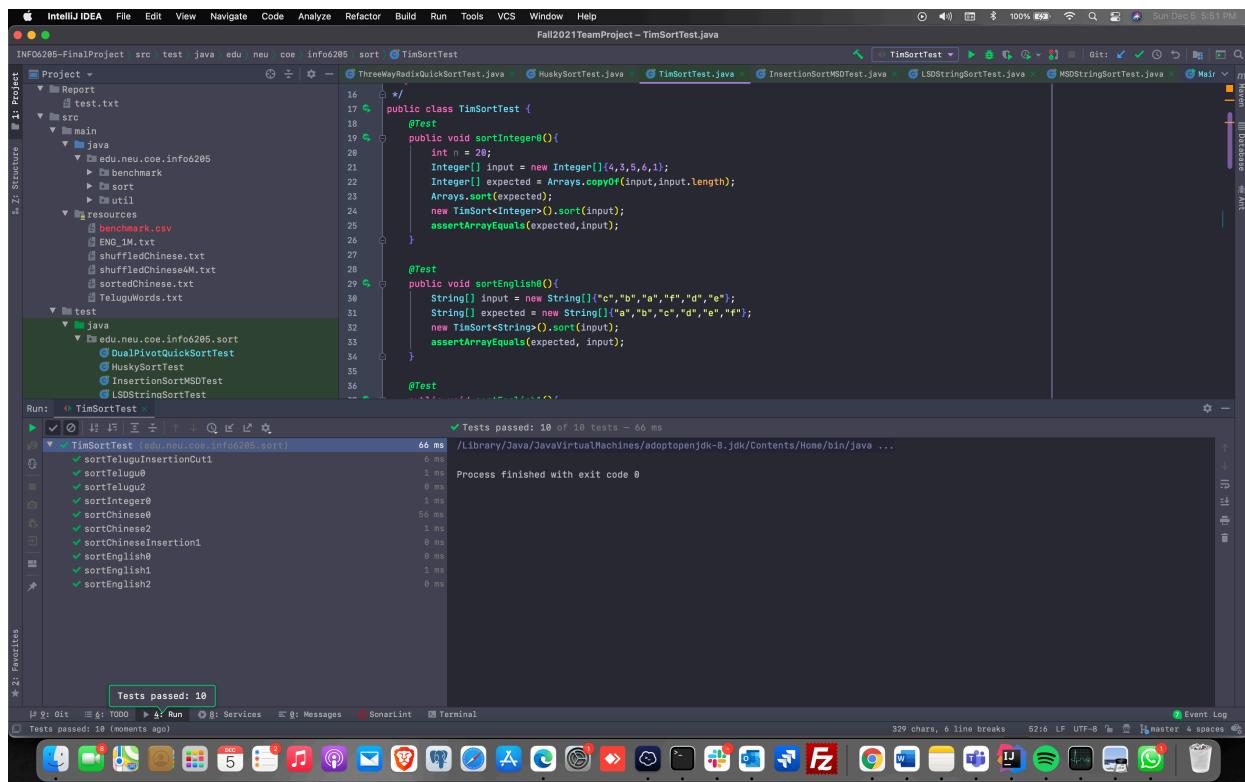
    @Test
    public void sortEnglish0() {
        String[] input = new String[]{"c", "b", "a", "f", "d", "e"};
        String[] expected = new String[]{"a", "b", "c", "d", "e", "f"};
        new TimSort<String>().sort(input);
        assertEquals(expected, input);
    }
}
```

Run: TimSortTest

Tests passed: 10 of 10 tests - 66 ms

Process finished with exit code 0

Tests passed: 10 (moments ago)



IntelliJ IDEA 2021.3.1

Fall2021TeamProject - ThreeWayRadixQuickSortTest.java

```
INFO6205-FinalProject src test java edu neu coe info6205 sort ThreWayRadixQuickSortTest.java HuskySortTest.java TimSortTest.java InsertionSortMSDTest.java LSDStringSortTest.java MSDStringSortTest.java
  Project Structure
  Screenshot
  Run: ThreWayRadixQuickSortTest
  Tests passed: 9
```

ThreeWayRadixQuickSortTest.java

```
package edu.neu.coe.info6205.sort;

import ...

public class ThreWayRadixQuickSortTest {

    @Test
    public void sortEnglish0() {
        String[] input = new String[]{"c", "b", "a", "f", "d", "e"};
        String[] expected = new String[]{"a", "b", "c", "d", "e", "f"};
        new ThreWayRadixQuickSort<String>().sort(input);
        assertEquals(expected, input);
    }

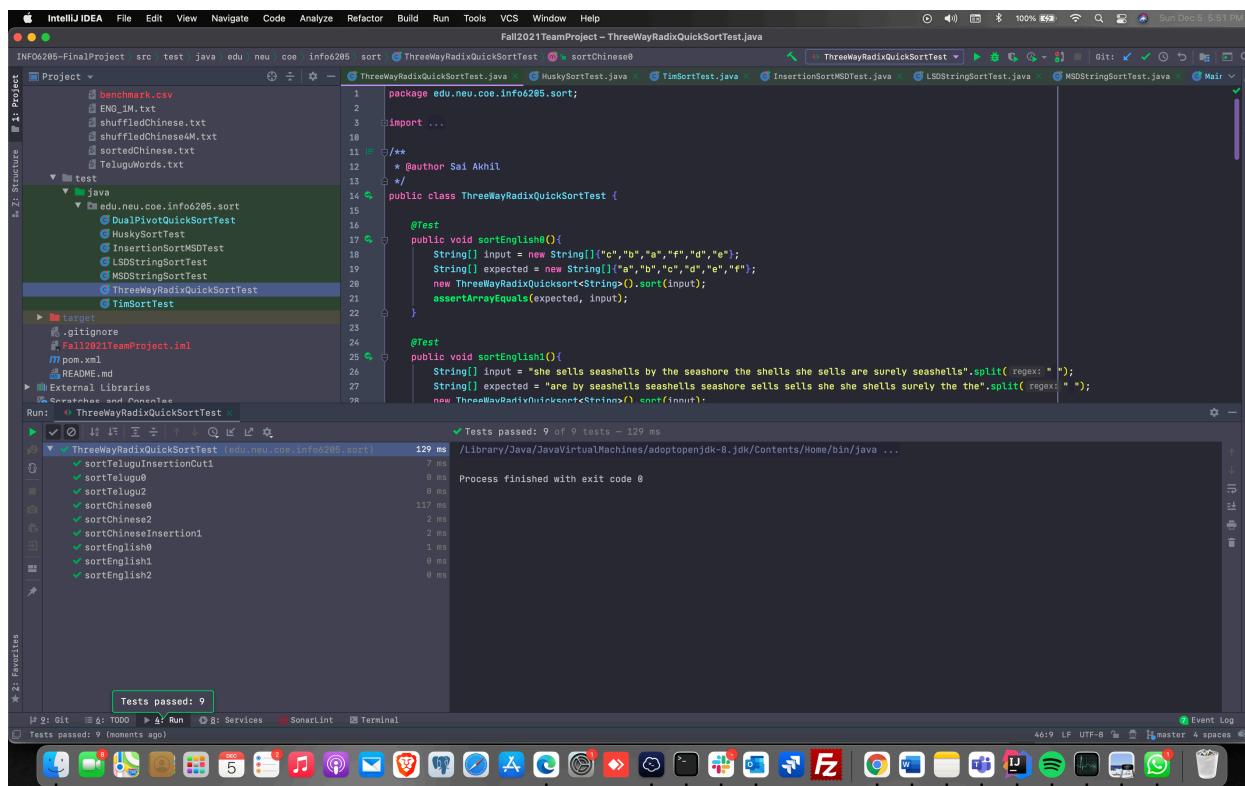
    @Test
    public void sortEnglish1() {
        String[] input = "she sells shells by the seashore shells she sells are surely shells".split(" ");
        String[] expected = "are by shells shells seashore sells sells she she shells surely the the".split(" ");
        new ThreWayRadixQuickSort<String>().sort(input);
        assertEquals(expected, input);
    }
}
```

Run: ThreWayRadixQuickSortTest

Tests passed: 9 of 9 tests - 129 ms

Process finished with exit code 0

Tests passed: 9 (moments ago)



IntelliJ IDEA 2021.3.3

Fall2021TeamProject - MSDStringSortTest.java

```
INFO6205-FinalProject src test java edu neu coe info6205 sort ThreewayRadixQuickSortTest.java HuskySortTest.java TimSortTest.java InsertionSortMSDTest.java LSDStringSortTest.java MSDStringSortTest.java
```

Project Structure

src

- edu.neu.coe.info6205.sort
- java
- test
- target
- .gitignore
- Fall2021TeamProject.iml
- pom.xml
- README.md
- External Libraries
- Scratches and Consoles

Run: MSDStringSortTest

Tests passed: 9 of 9 tests - 231 ms

MSDStringSortTest (edu.neu.coe.info6205.sort)

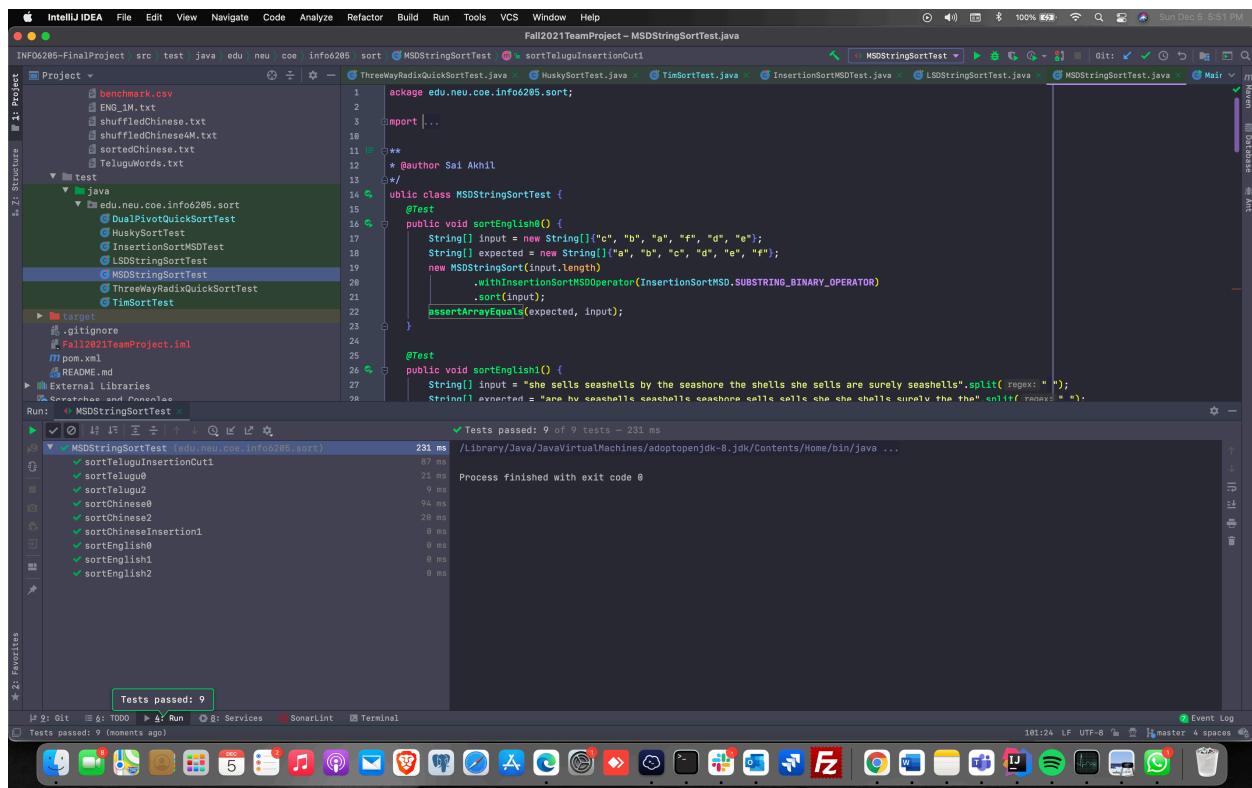
- sortTeluguInsertionCut1
- sortTelugu0
- sortTelugu1
- sortTelugu2
- sortChinese0
- sortChinese1
- sortChinese2
- sortChineseInsertion1
- sortEnglish0
- sortEnglish1
- sortEnglish2

Process finished with exit code 0

Tests passed: 9 (moments ago)

Event Log

101:24 LF UTF-8 master 4 spaces



IntelliJ IDEA 2021.3.3

Fall2021TeamProject - LSDStringSortTest.java

```
INFO6205-FinalProject src test java edu neu coe info6205 sort ThreewayRadixQuickSortTest.java HuskySortTest.java TimSortTest.java InsertionSortMSDTest.java LSDStringSortTest.java MSDStringSortTest.java LSDStringSortTest.java
```

Project Structure

src

- edu.neu.coe.info6205.sort
- java
- test
- target
- .gitignore
- Fall2021TeamProject.iml
- pom.xml
- README.md
- External Libraries
- Scratches and Consoles

Run: LSDStringSortTest

Tests passed: 9 of 9 tests - 177 ms

LSDStringSortTest (edu.neu.coe.info6205.sort)

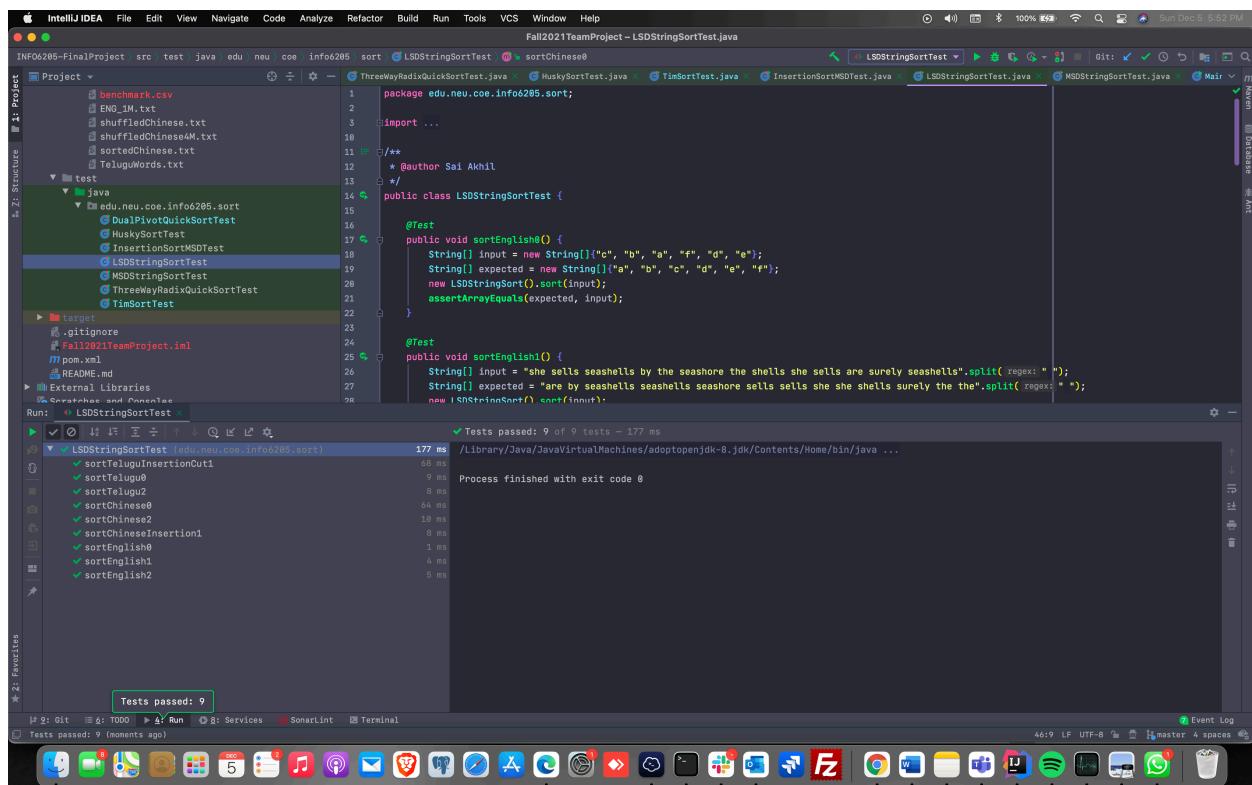
- sortTeluguInsertionCut1
- sortTelugu0
- sortTelugu1
- sortTelugu2
- sortChinese0
- sortChinese1
- sortChinese2
- sortChineseInsertion1
- sortEnglish0
- sortEnglish1
- sortEnglish2

Process finished with exit code 0

Tests passed: 9 (moments ago)

Event Log

46:9 LF UTF-8 master 4 spaces



IntelliJ IDEA

Fall2021TeamProject - InsertionSortMSDTest.java

```
package edu.neu.coe.info6205.sort;

import ...  
import java.util.Arrays;  
import java.util.Collections;  
  
public class InsertionSortMSDTest {  
  
    @Test  
    public void sortInteger1() {  
        Integer[] expected = new Integer[]{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16};  
        Integer[] input = Arrays.copyOf(expected, expected.length);  
        Collections.shuffle(Arrays.asList(input));  
        new InsertionSortMSDInInteger().sort(input, position: 0);  
        assertEquals(expected, input);  
    }  
  
    @Test  
    public void sortInteger0() {  
        Integer[] input = new Integer[]{4, 3, 5, 1, 2};  
    }  
}
```

Tests passed: 11

Run: InsertionSortMSDTest

InsertionSortMSDTest (edu.neu.coe.info6205.sort) 170 ms /Library/Java/JavaVirtualMachines/adoptopenjdk-8.jdk/Contents/Home/bin/java ...
sortTeluguInsertionOut1 81 ms
sortTelugu0 2 ms
sortTelugu2 1 ms
sortInteger8 2 ms
sortInteger1 1 ms
sortChinese9 75 ms
sortChinese2 4 ms
sortChineseInsertion1 1 ms
sortEnglishH 1 ms
sortEnglishM1 1 ms
sortEnglish2 1 ms

Process finished with exit code 0

Tests passed: 11 (moments ago)

Event Log

IntelliJ IDEA

Fall2021TeamProject - HuskySortTest.java

```
package edu.neu.coe.info6205.sort;

import edu.neu.coe.info6205.sort.huskySort.HuskyCoderFactory;
import edu.neu.coe.info6205.sort.huskySort.PureHuskySort;
import org.junit.Test;

import java.util.Arrays;
import java.util.Collections;
import static org.junit.Assert.assertArrayEquals;

public class HuskySortTest {
    @Test
    public void sortChinese99() {
        String[] expected = new String[]{"同安", "阿彬", "阿兵", "阿冰", "阿晨", "同称", "阿丛", "阿丹", "阿德力", "阿迪迪", "阿江", "阿迪雅", "阿栋", "阿繁", "阿方", "阿非", "阿八"};
        String[] input = Arrays.copyOf(expected, expected.length);
        Collections.shuffle(Arrays.asList(input));
        new PureHuskySort<>(HuskyCoderFactory.asciiCoder, mayBeSorted: false, useInsertionSort: false).sort(input);
        assertArrayEquals(expected, input);
    }
}
```

Tests passed: 3

Run: HuskySortTest

HuskySortTest (edu.neu.coe.info6205.sort) 112 ms /Library/Java/JavaVirtualMachines/adoptopenjdk-8.jdk/Contents/Home/bin/java ...
sortChinese99 110 ms
sortChinese98 1 ms
sortChinese99 1 ms

Process finished with exit code 0

Tests passed: 3 (moments ago)

Event Log

IntelliJ IDEA 2021.3.1

Fall2021TeamProject - DualPivotQuickSortTest

Project Structure

src test java edu neu coe info6205 sort

DualPivotQuickSortTest

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help

Code Editor (DualPivotQuickSortTest.java)

```
1 package edu.neu.coe.info6205.sort;
2
3 import java.io.FileReader;
4 import org.junit.Test;
5
6 import java.io.IOException;
7 import java.util.Arrays;
8 import java.util.Collections;
9 import java.util.Locale;
10
11 import static org.junit.Assert.assertArrayEquals;
12
13 /**
14 * @author Sai Akhil
15 */
16
17 public class DualPivotQuickSortTest {
18
19     @Test
20     public void sortInteger() {
21         Integer[] input = new Integer[]{4, 3, 5, 6, 1};
22         Integer[] expected = Arrays.copyOf(input, input.length);
23
24         sortTeluguInsertionOut();
25         sortTelugu0();
26         sortTelugu2();
27         sortInteger0();
28         sortInteger1();
29         sortChinese0();
30         sortChinese2();
31         sortChineseInsertion1();
32         sortEnglish0();
33         sortEnglish1();
34         sortEnglish2();
35
36         assertArrayEquals(expected, input);
37     }
38 }
```

Run: DualPivotQuickSortTest

Tests passed: 11 of 11 tests - 107 ms

sortTeluguInsertionOut 107 ms /Library/Java/JavaVirtualMachines/adoptopenjdk-8.jdk/Contents/Home/bin/java ...

sortTelugu0 0 ms

sortTelugu2 0 ms

sortInteger0 0 ms

sortInteger1 1 ms

sortChinese0 91 ms

sortChinese2 2 ms

sortChineseInsertion1 1 ms

sortEnglish0 1 ms

sortEnglish1 1 ms

sortEnglish2 0 ms

Tests passed: 11

Event Log

340 chars, 6 line breaks 61:6 LF UTF-8

Icons: Git, TODO, Run, Services, SonarLint, Terminal