Peixin Yao(001526352)

Rongjin Dang (001565800)

Zixuan Zhu (002196924)

**Program Structures & Algorithms**

**Fall 2021**

**Final**

* **Task (List down the tasks performed in the Assignment)**

Your task is to implement MSD radix sort for a natural language which uses Unicode characters. You may choose your own language or (Simplified) Chinese. Additionally, you will complete a literature survey of relevant papers and you will compare your method with Timsort, Dual-pivot, uicksort, Huskysort, and LSD radix sort.

* **How to implement Msd**

1: we transform Chinese to pinyin(with tone)

2: we add Unicode(UTF-8) to pinyin.(Some Chinese have the same pinyin and same tone. So when pinyin and tone can’t distinguish them, we use Unicode to sort them)

3: we use map(or aux array) to record Chinese.

4: we use Msd to sort pinyin, and when lo<=hi+15 we use insertion sort

5: we get chinese from pinyin, there are two methods. First method, We get their Corresponding Chinese characters from the map based on a sorted array of pinyin. The other one, when we change pinyin array, we also do the same change for the aux array recording Chinese. Finally, we find using aux array is more efficient.

* **Relationship Conclusion: (For ex : z = a \* b)**

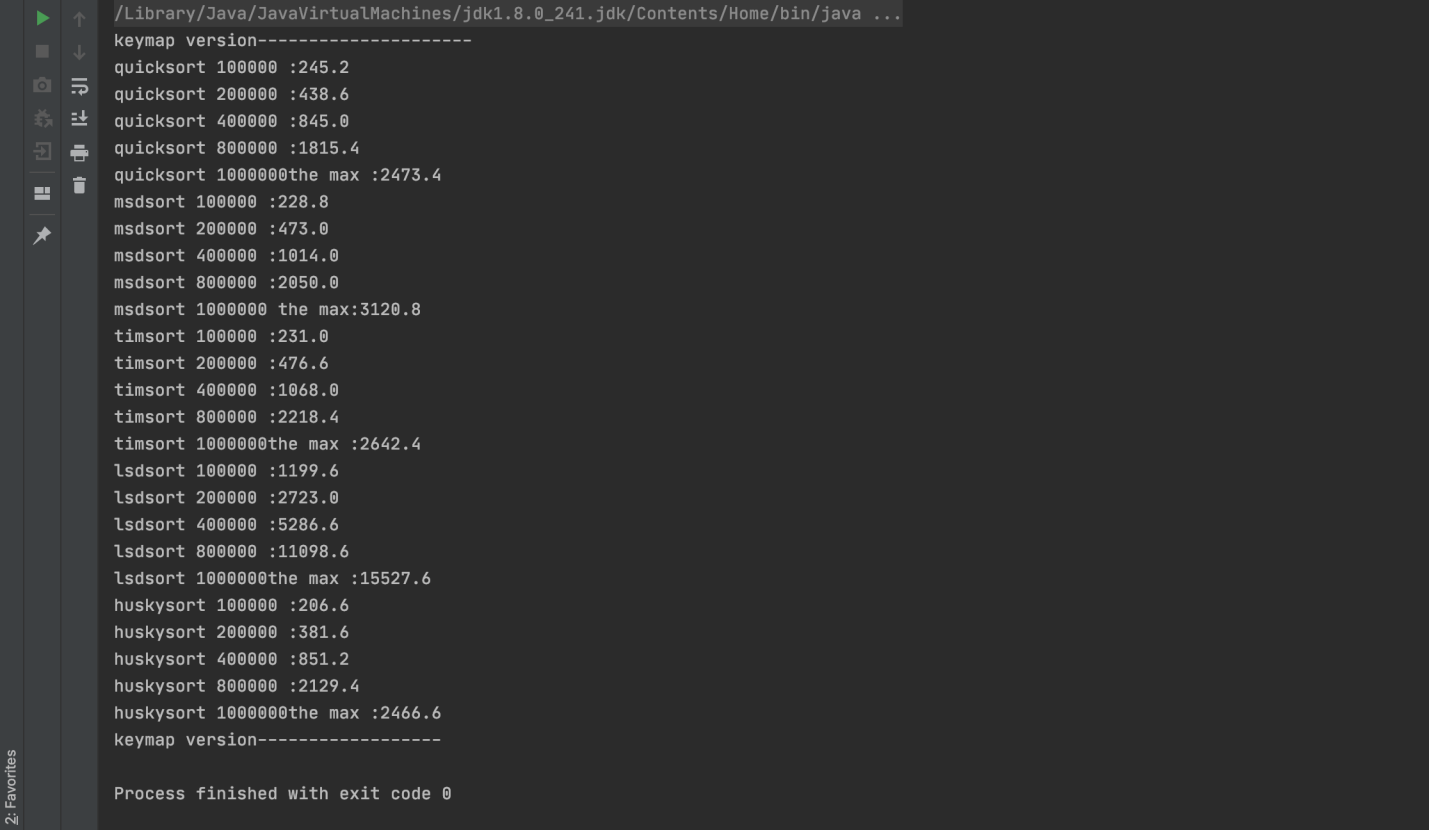
**From The results of Benchmark, we finds that the PureHuskySort is the Fastest algorithm, we recommend using PureHuskySort to sort Chinese.**

**MSD radix sort and LSD radix sort also have obvious advantages. when string is short, the efficiency is very good. For example, when we sort an array that every element in array only have one or two length, the speed of MSD is eight times faster than quicksort. But we transform Chinese to pinyin with tone, most of the pinyin length have already exceeded over 10. In order to distinguish them when they have the same pinyin and tone, we add Unicode to them. The length of them is long. In this Case, comparison sort is better. Therefore, I commend husky sort.**

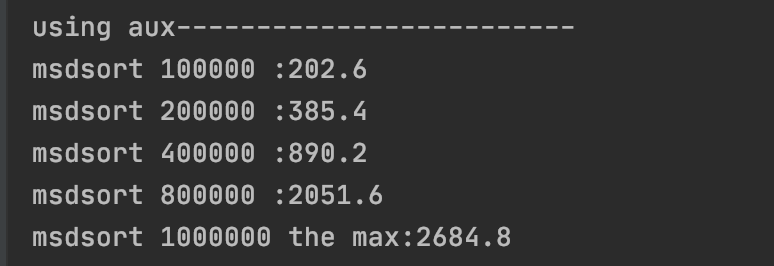
* **Evidence to support the conclusion:**

1. **Output (Snapshot of Code output in the terminal)**

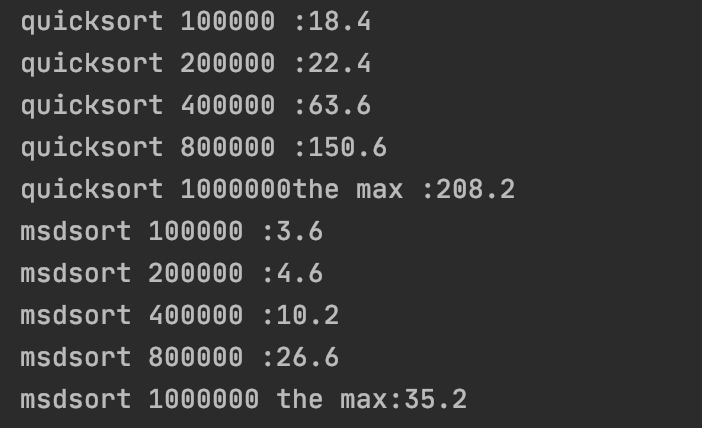
**MSD radix sort, Dual-pivot Quicksort, Timsort, Huskysort and LSD radix sort (use map)**



**MSD radix sort(use aux array)**

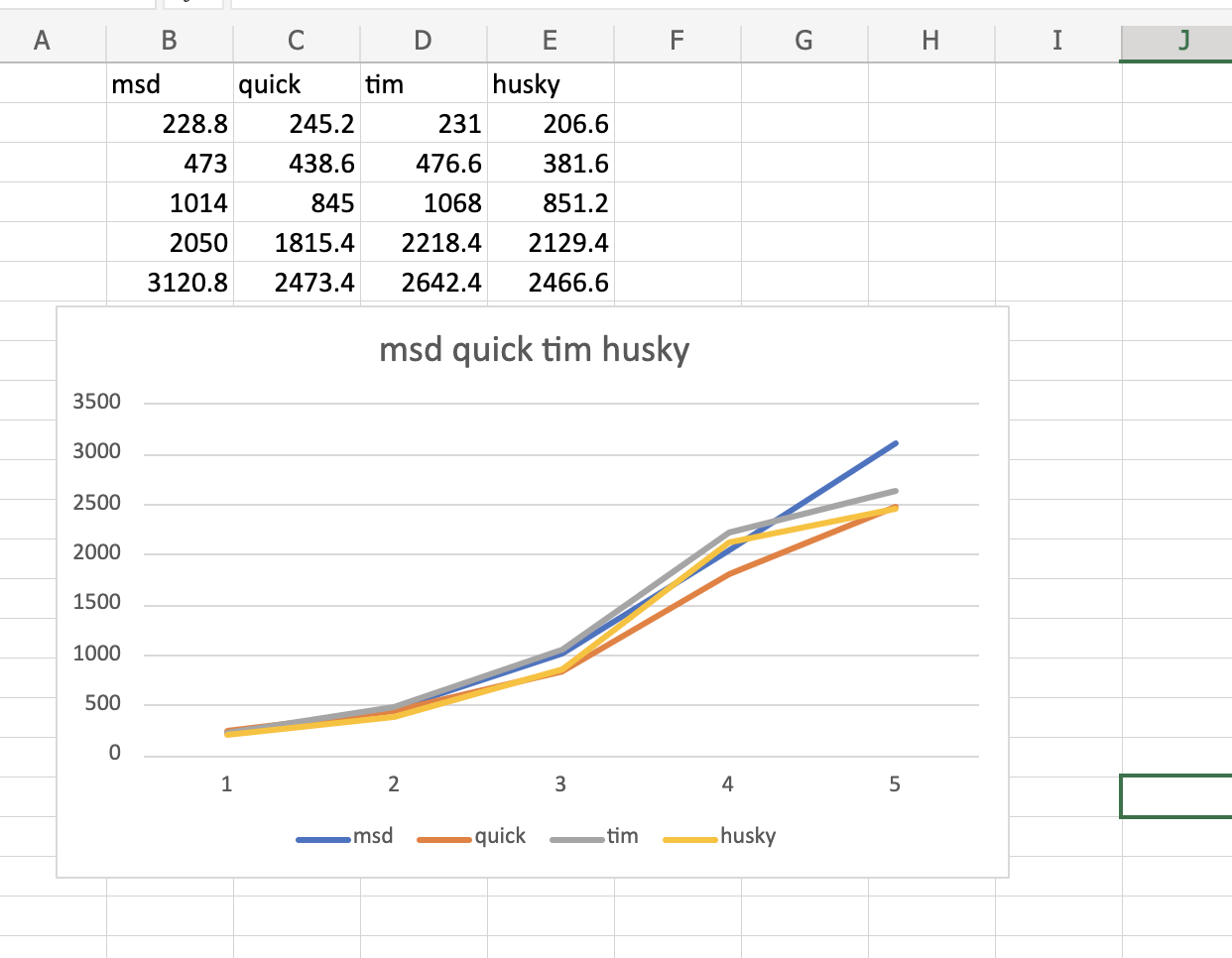


**Compare MSD and quickSort when every element in array only have one length**

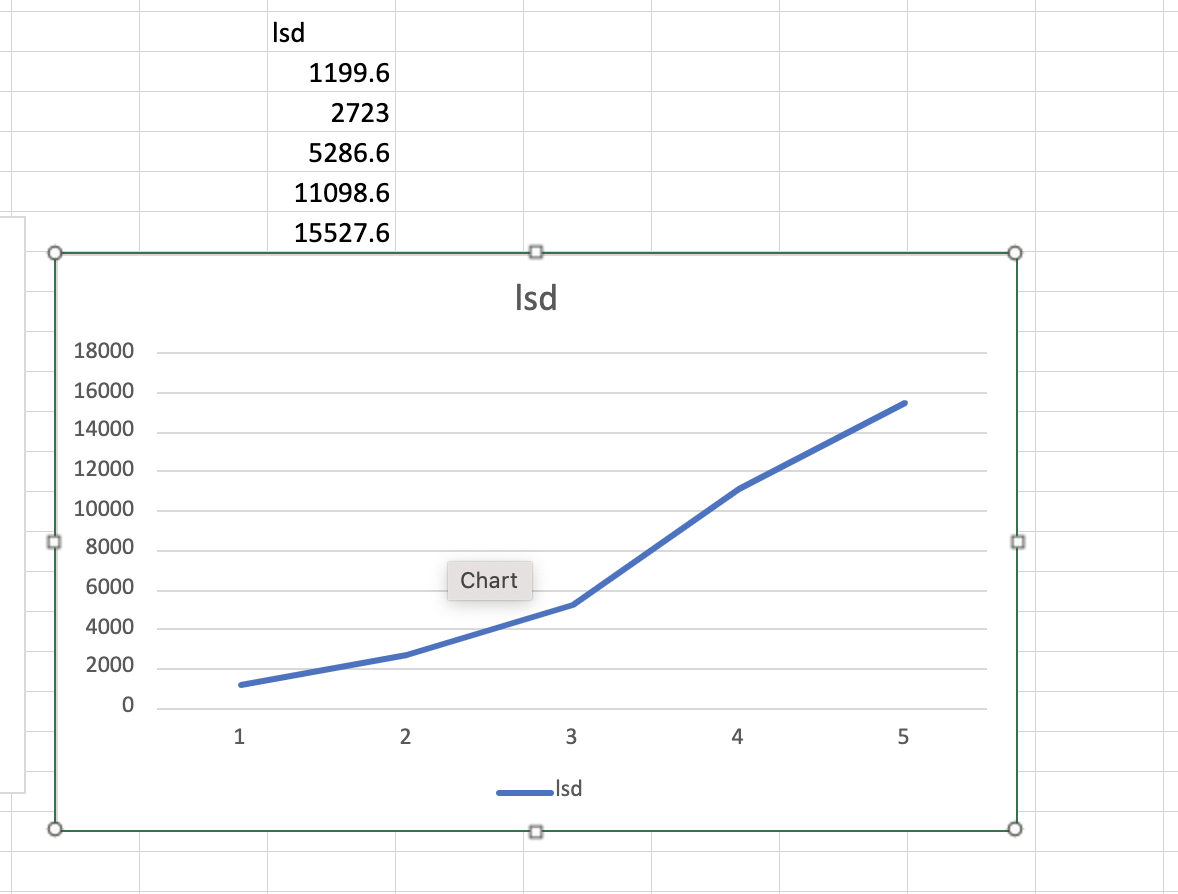


1. **Graphical Representation(Observations from experiments should be tabulated and analyzed by plotting graphs(usually in excel) to arrive on the relationship conclusion)**

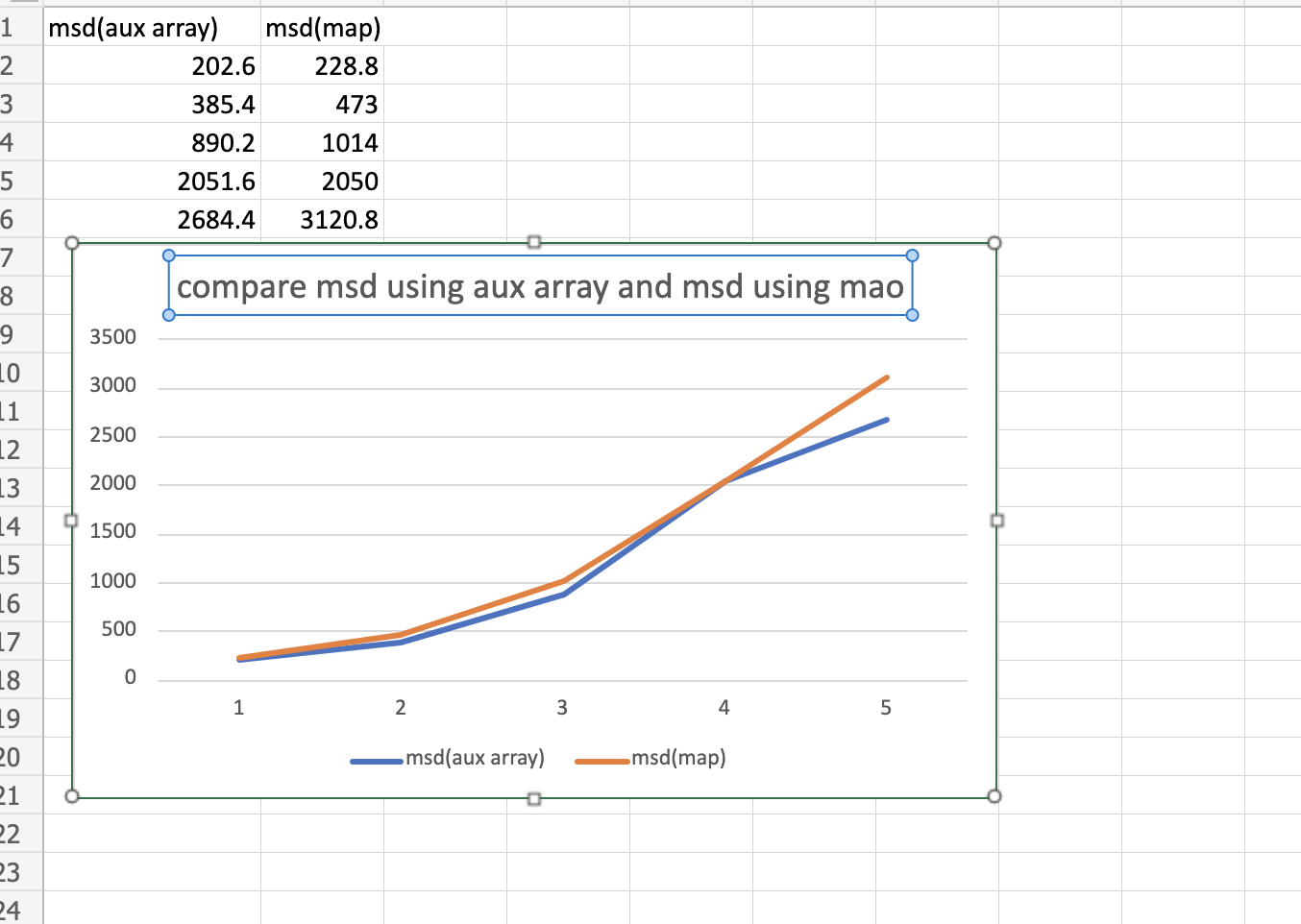
**MSD radix sort, Dual-pivot Quicksort, Timsort, Huskysort(use map)**



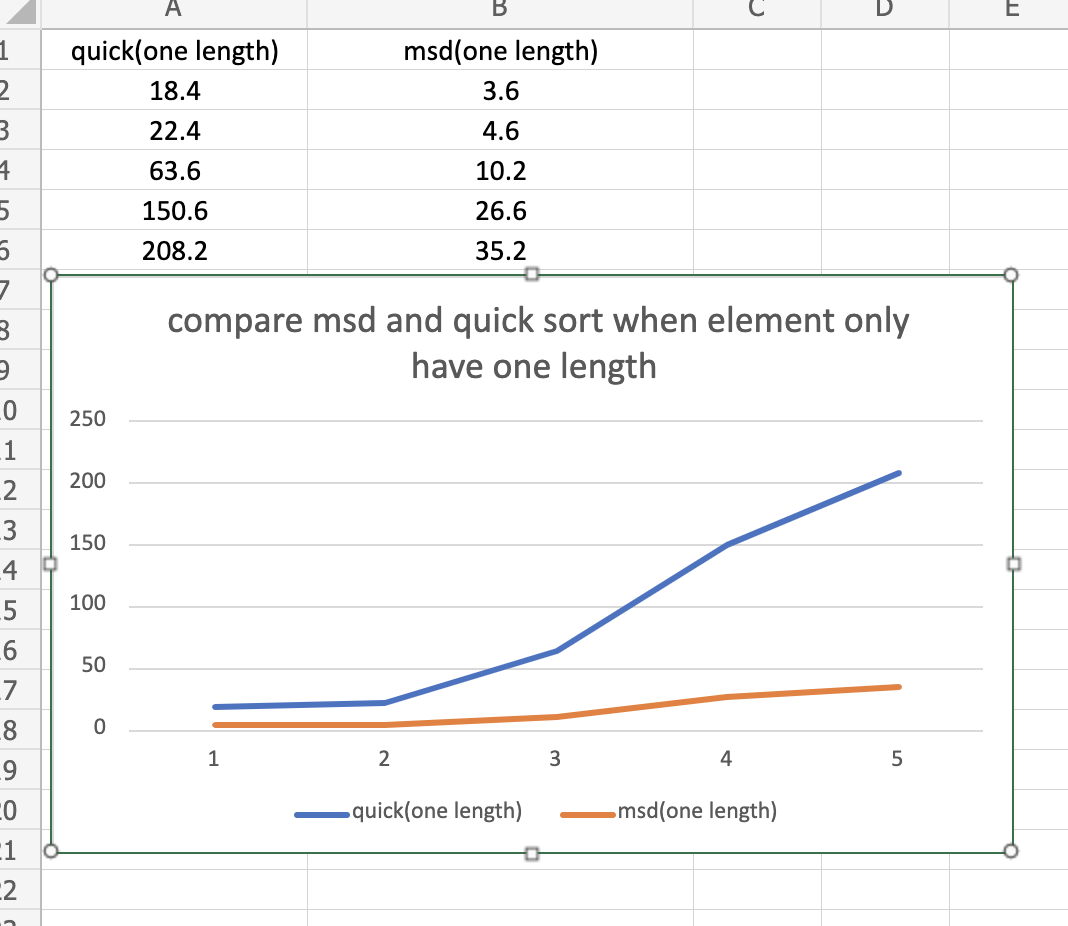
**LSD radix sort(use map)**



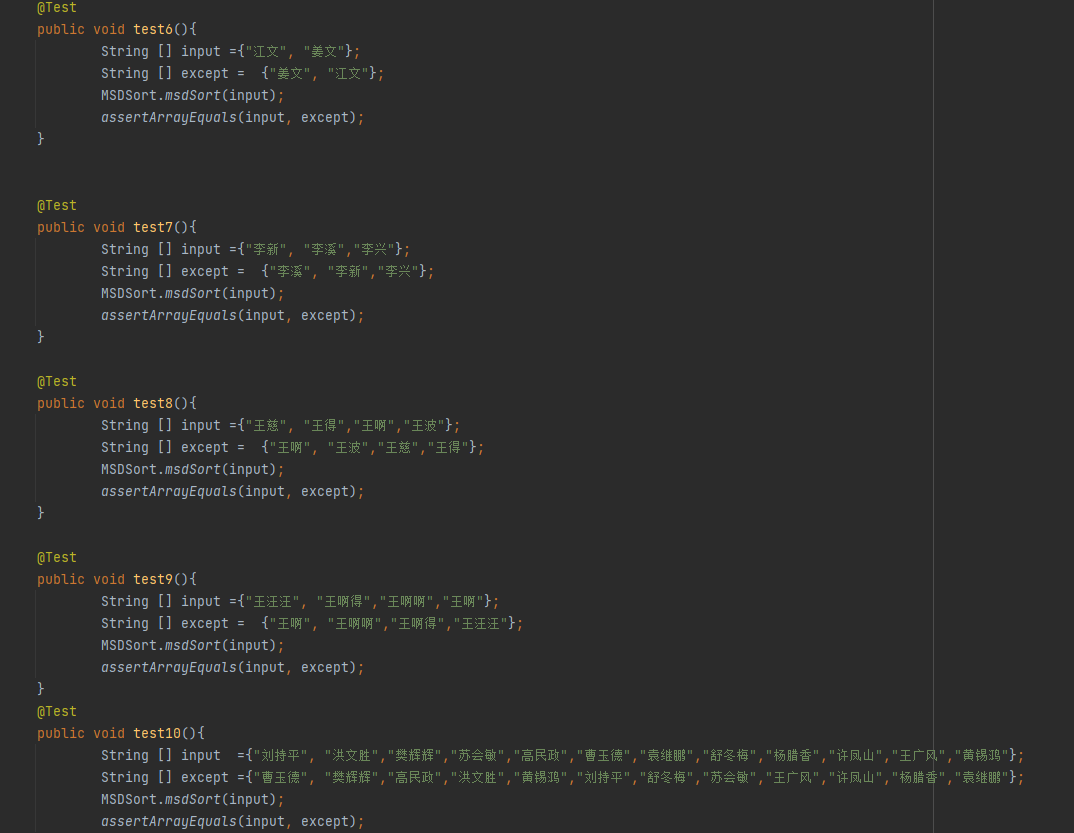
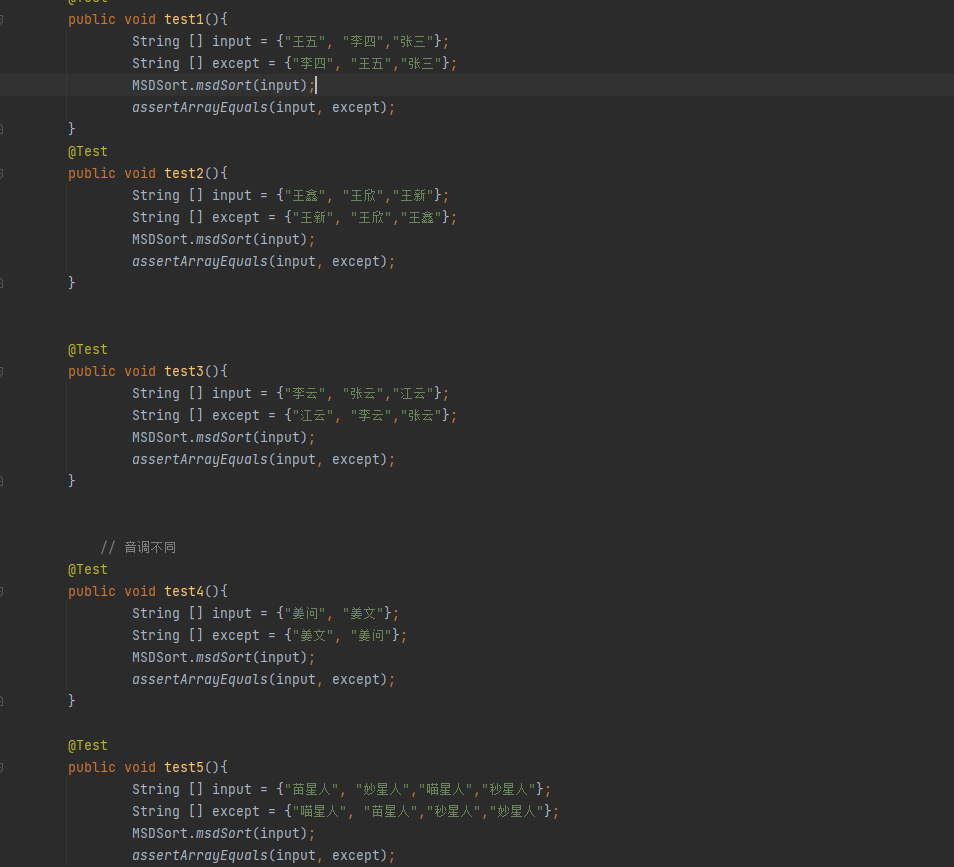
**compare msd using aux array and msd using map**



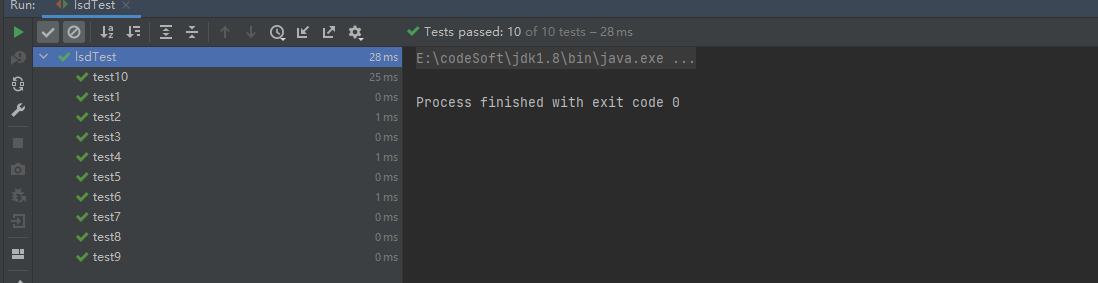
**compare msd and quick sort when element only have one length**



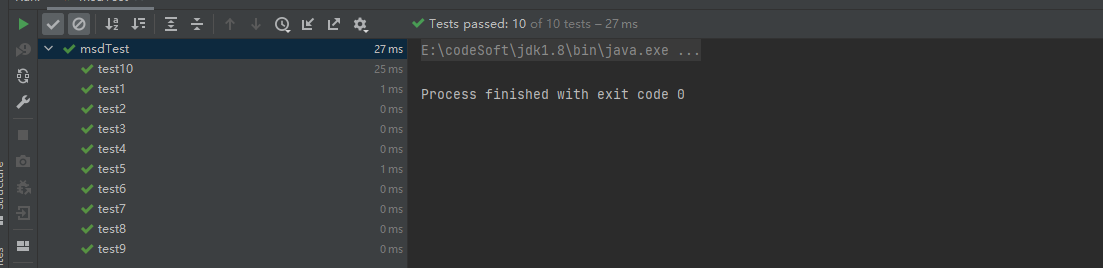
* **Unit tests result:(Snapshot of successful unit test run)**



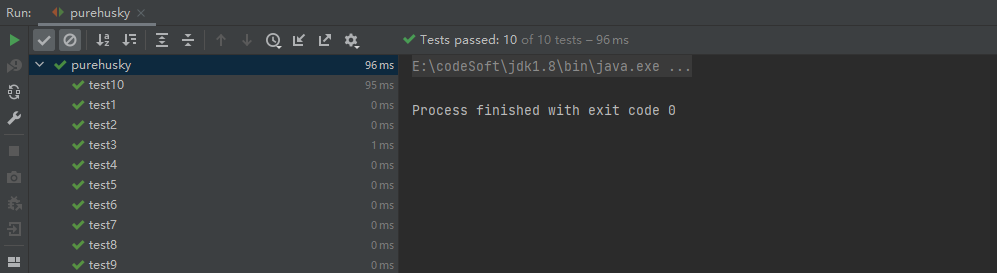
LSD Test



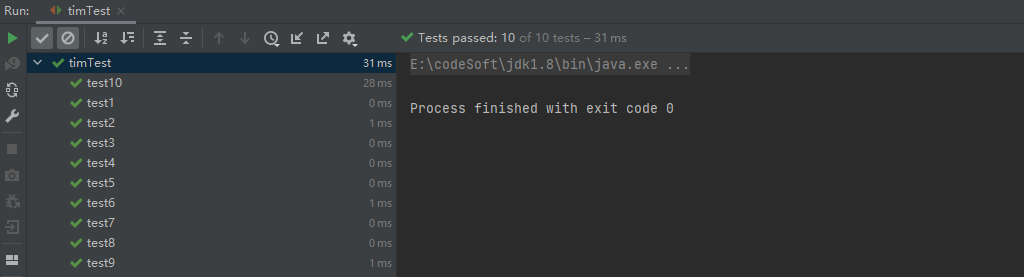
MSD Test



Husky Test



tim Test



Quick Test

