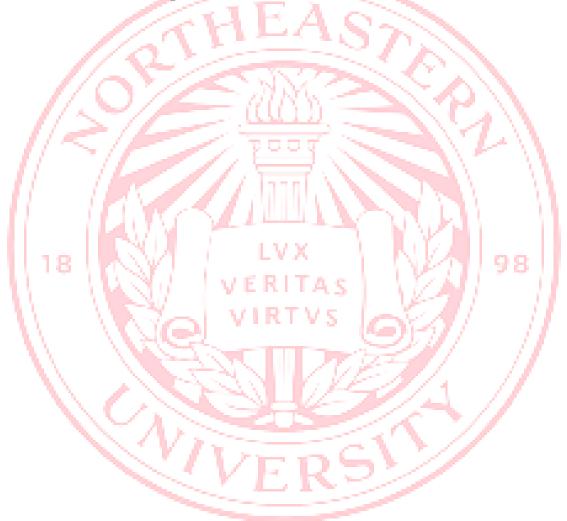
# A Dissertation Presented By Yanrong Huang 001563514 Yuan Lian 001063563 Yuxin Liu 001028722

PROGRAM STRUCTURES & ALGORITHMSFALL 2021 FINAL

# Contents

1	Task	1
2	How we Realize	2
3	Output	3
4	Unit tests result : snapshot of successful unit test run	4



### Task

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 Quicksort, Huskysort, and LSD radix sort.

### How we realize

### 1. Trans Chinese to Pinyin

Chinese characters can be stored as a unicode string or as a char. In order to sort Chinese characters in pronunciation order, all characters must be transferred to sortable Pinyins. Therefore, we implement five functions: 1) unicodeToChar, 2) charToPinyinNOTone, 3) charToPinyinWITone, 4) unicodeToPinyinNOTone, and 5) unicodeToPinyinWITone.

# 2.Implement Different Sorting Algorithm

We used Prof Robin's Timsort, pureHusky sort ,LSD radix sort and we rewrite the MSD radix sort and Dual-pivot Quicksort. The MSD radix sort comes from geeksforgeeks and the Dual-pivot Quicksort is from Algorism 4<sup>th</sup>.

### 3.Benchmark Timer

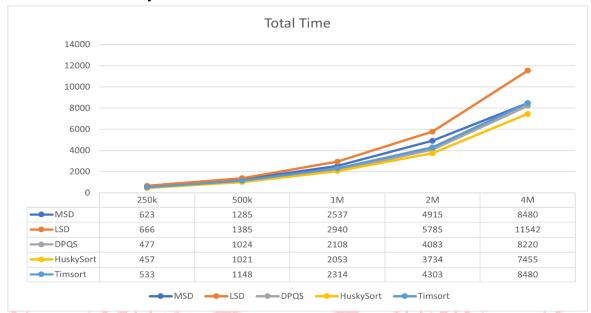
We rewrite the timer method and name it Timer. We record the sorting start system time and then record the sorting end system. And we minus this two times to count the total time of our sorting algorism. But this method can only use when the input is large. Since we count as microseconds.

### 4. The whole process

We first read the input method from shuffledChinese.txt and then we convert it to pinyin. We copied this pinyin string and sort it. Finally, we use matchIndex() methods to match the origin Chinese string, pinyin String and sorted pinying string. And finally we use file writer to print output.

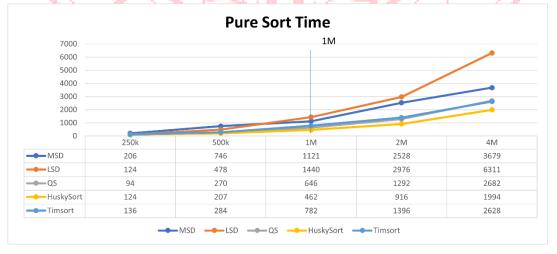
# Output

1. The following is a screenshot of the results of multiple algorithms running on different scales of data and only count the **total sort time**.

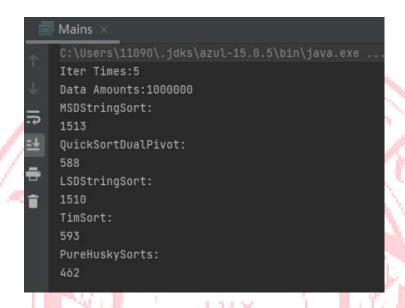


This picture is a line graph display of the running time of the same algorithm on data of different scales. See from this picture, we find that Husky sort algorithm is the most efficient algorism. The growth trend among them is similar. Dual-pivot Quicksort and Tim sort have the same speed in most of time. LSD performs better on small-scale data, but on larger data volumes, time consumption is greatly increased. The MSD algorithm performs poorly on small-scale data, but performs better on larger-scale data.

The following is a screenshot of the results of multiple algorithms running on different scales of data and only count the **pure sort time**.



2. Meanwhile, based on 1M input as we have seen, the most efficient algorithm is the HuskySort algorithm, which runs less than the other four algorithms on the same amount of data. And at this time MSD and LSD are slowly .Dual-Pivot quicksort and Tim Sort are as quick as each other.



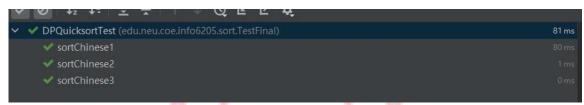
3. We also print the output as a txt file when input is 1M data.

■ ary.txt - 记事本 文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

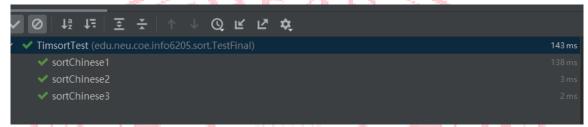
阿安,阿斌,阿斌,阿斌,阿冰,阿冰,阿冰,阿滩,阿蝉,阿朝,阿朝,阿臣,阿臣,阿臣,阿臣,阿诚,阿诚,阿诚,阿诚,阿弛,阿弛, ,艾恒,艾和平,艾宏,艾宏,艾宏,艾宏,艾宏,艾宏,艾红兵,艾洪波,艾洪波,艾洪德,艾红光,艾宏国,艾红红,艾红娟 爱梅,爱梅,爱梅,爱梅,爱梅,灵美儿,艾美华,艾美君,艾梅兰,艾美丽,艾美林,艾美玲,艾美美,艾美娜,艾美琪,艾美 娜,艾伟平,艾伟平,艾微微,艾微微,艾微微,艾微微,艾微微,艾微微,艾文,艾文,艾文,艾文,艾文,艾文妹,艾文兵,艾文博, 平,艾玉琴,艾玉清,艾玉菜,艾玉文,艾玉霞,艾玉珍,艾玉洲,艾泽,艾扎,艾长春,艾长虎,艾长华,艾长明,艾长青,艾 春红,安春华,安春辉,安春兰,安春雷,安春丽,安春霖,安春玲,安春龙,安春梅,安春明,安春山,安春生,安春霞,安 政,安国忠,安国柱,安海,安海波,安海峰,安海峰,安海红,安海华,安海娟,安海军,安海利,安海利,安海林,安海龙 安静会,安菁菁,安菁菁,安菁菁,安菁菁,安菁美,安静美,安静敏,安敬宁,安静平,安静茹,安静生,安劲松,安静涛,安静伟, 汀,安美霞,安美燕,安美伊,安美英,安萌,安萌,安萌,安萌,安蒙蒙,安梦娜,安梦琪,安梦婷,安梦瑶,安弥,安弥,安 师,安师,安师,安师,安师,安师,安师,安士波,安世昌,安世超,安士城,安世东,安世峰,安世高,安世耿,安世豪,安 ,安晓慧,安晓杰,安晓静,安晓娟,安晓军,安晓坤,安晓兰,安晓乐,安晓丽,安晓亮,安孝林,安孝林,安孝林,安晓玲 安仪,安仪,安仪,安仪,安仪,安仪,安一凡,安一方,安逸飞,安沂华,安奕霖,安一敏,安一敏,安银,安银,安银,安银 ,安志强,安志儒,安芷生,安志伟,安志文,安志信,安治星,安志学,安芷瑶,安志义,安志英,安志勇,安志远,安志远, 双英,敖顺,傲丝,敖四海,敖思红,敖思祥,傲松,敖松林,奥陶,傲特,傲腾,傲腾,傲腾,敖天,奥婷,敖婷婷,敖通,敖3 ,阿悦,阿悦,阿悦,阿云,阿云锐,阿湛,阿湛,阿昭,阿震,阿震,阿震,阿震,阿震,阿震,阿震,阿政,阿政,阿治,阿治, ,白春龙,白春茂,白春梅,白春敏,白春明,白春鹏,白春萍,白春萍,白春仁,白春生,白春堂,白春霞,白春香,白春香 阁,白根,白庚,白庚胜,白庚延,白公胜,白谷,白冠,白广,白广,白广才,白光大,白广东,白光光,白广海,白光虎,白 杰,白红菊,白红娟,白红军,白红军,白红兰,白宏丽,白宏丽,白宏丽,白红亮,白红亮,白红亮,白红梅,白红梅,白红梅 ,白嘉轩,白佳莹,白嘉雨,白嘉雨,白佳媛,白佳媛,白继超,白继承,白继东,白洁,白洁,白洁,白洁,白洁,白洁,白洁,白洁冰, ,白俊涛,白俊文,白俊武,白俊艳,白俊艳,白俊毅,白俊毅,白俊毅,白俊英,白驹荣,白巨星,百卡,白凯,白开军 龙江,白龙驹,白龙龙,白龙强,白龙水,白龙玉,白龙珠,白陆,白陆,白陆,白陆,白陆,白陆,白陆,白陆,白强、白鹭飞, ,白庆辉,白清江,白庆军,白清兰,白清兰,白庆利,白清亮,白清林,白清林,白清林,白清林,白庆玲,白庆龙,白清明 利,白胜男,白生鹏,白生生,白胜文,白生银,白师,白师,白师,白师,白世彬,白世彬,白世昌,白世超,白世超,白世兄 白彤,白彤,白彤,白彤,白彤,白彤东,白同平,白同朔,白彤彤,白图,白团,白拓,白托娅,白图雅,白万,白万,白万,白 ,白晓芬,白晓锋,白晓锋,白晓锋,白晓刚,白晓光,白晓航,白晓虹,白晓虹,白晓华,白晓慧,白晓慧,白晓佳,白晓剑 姐,白雪静,白雪静,白学军,白雪来,白雪雷,白雪利,白雪利,白雪莲,白雪亮,白雪亮,白雪林,白雪林,白雪林,白雪 白亚平,白雅琪,白雅琴,白雅琴,白亚青,白亚茄,白亚涛,白雅婷,白亚伟,白亚文,白亚文,白亚文,白雅欣,白雅欣, 白渊,白渊,白渊,白渊,白渊,白洪,白元光,白元明,白元明,白园园,白园园,白园园,白园园,白园园,白玉宝,白玉宾,白玉 芸,白云章,白云珍,白云志,白云中,白云子,白云子,白云宗,白玉佩,白羽鹏,白羽鹏,白玉平,白玉平,白玉平,白玉平, .白志勋,白志艺,白志革,白志革,白智重,白智重,白志强,白志忠,百众,百众,百众,百众,白忠宝,白忠德,白重恩

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

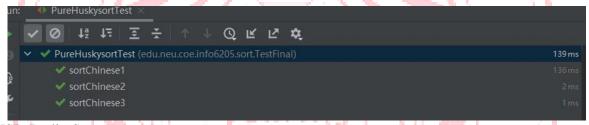
1. Dual-Pivot quicksort



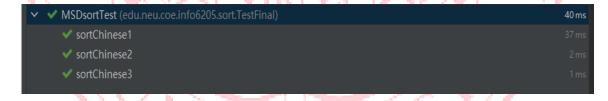
### 2.Tim Sort



# 3. PureHusky Sort



### 4.MSD Radix Sort



VIRTUS

### 5.LSD Radix Sort

