

Homework1

The result of both task1 and task2 is shown in the folder of *result*. My source code for task1 and task2 is called *ChiWei_Liu_task1.py* and *ChiWei_Liu_task2.py*. Also, the source data of small dataset and big data set is put in the folder of *small_data* and *big_data* separately.

How to run the task 1:

- For small data set: Open the source code *ChiWei_Liu_task1.py* and go to line 6 of the code, change the value of *sourceFold* to “small_data”, which indicate the folder of input data source, and you can change the name of the out put file by go to line 7 of the code, change the value of *Outputfile*. Then run the code, and you will get the result at the *result* folder.
- For big data set: Open the source code *ChiWei_Liu_task1.py* and go to line 6 of the code, change the value of *sourceFold* to “big_data”, which indicate the folder of input data source, and you can change the name of the out put file by go to line 7 of the code, change the value of *Outputfile*. Then run the code, and you will get the result at the *result* folder.

How to run the task 2:

- For small data set: Open the source code *ChiWei_Liu_task2.py* and go to line 6 of the code, change the value of *sourceFold* to “small_data”, which indicate the folder of input data source, and you can change the name of the out put file by go to line 7 of the code, change the value of *Outputfile*. Then run the code, and you will get the result at the *result* folder.
- For big data set: Open the source code *ChiWei_Liu_task2.py* and go to line 6 of the code, change the value of *sourceFold* to “big_data”, which indicate the folder of input data source, and you can change the name of the out put file by go to line 7 of the code, change the value of *Outputfile*. Then run the code, and you will get the result at the *result* folder.