

Dingbang Chen
Joshua Li
9/28/17
Professor Tjang
CS214 Project 0

Readme

We used a triple pointer to store the dataset, and a string array to store the column titles. At first, we used strtok to tokenize all the data in csv file. However, we found out that there are empty columns in the file. It may cause problems when we tokenize the data. So, we used strtok to tokenize the column titles. For the data part, we put the code into a method called readdata() for clarity, its prototype is included in the header file. We used getchar() to see if there is an empty column. Another advantage of using getchar() is that we can identify if there is a comma in the data which has quotation marks. After this, we found out we need to skip leading/trailing spaces as well, and we fixed it. We set up a maximum number of columns and lines. If the data reaches the maximum number, our program will realloc space to prevent errors. When executing sorter, it is important to identify if the column is a number or a string. So, in our program, if the column in the first line is detected to be a string, this column will be identified as a column with type string. If the first char is detected to be a number, the program will check the rest of the string to confirm the type of this column. We created a method called isnum(), and its prototype is also stored in the header file. Last but not least, we used recursive mergesort to sort the data.

Assumptions

A single column will not exceed 10000 characters.
A single line in csv file will not exceed 50000 characters.
There are no empty column titles.