

Bhavin Soni : 171002303
Kyle Chen : 163000952

Description:

=====

This is a multi-process C program to sort a large amount data, specifically a list of movie records from imdb, alphabetically by the data given in the specified column. The program will fork() a child process within a given directory and output the results in a different file. The program will output in the terminal the PID of the process and its child processes.

Usage:

=====

Program must take in the "-c" parameter to indicate the specific column to sort the data.

Example: ./sorter -c duration

The "-d" flag indicates the directory the program will search for .csv files

Example: ./sorter -c duration -d thisDirectory/thatDirectory

The "-o" flag indicates the output directory for the sorted versions of the input file

Example: ./sorter -c duration -d thisDirectory -o thatDirectory

Design: (Header)

=====

getthis_line = Reads the lines from the .csv file

str_token = own tokenizer function used to help separate the data from the csv file

mergesort, sort, merge = own mergesort function used to sort divide the csv data into arrays and merge them together sorted in alphabetical order

parseRowtoint = struct that takes the row values for each column so the data is put into the right columns/rows

Rowtypes/Columntypes = used to determine the data values for each column or row

printCSV = prints the data into a csv format

Sorter = function used to sort the csv files

Status = checks the directory to determine if the .csv file is sorted or not

direct = traverses through the directories and forks child processes on the working directories

Difficulties & Challenges:

=====

-Originally our mergesort would put spaces between each row in the csv file, but we were able to fix that. However, we could not get our mergesort to work as efficiently or accurately as we hoped - Our mergesort alphabetically sorts the data in different blocks - we were unable to fix our merge function

- The printing of the PIDs were kind of a small issue. The initial PID process will be printed out correctly. The PIDS of the child processes will be printed within the lines of "File is not a .csv".