PA-2 Description:

Note:

before running merge sort create two folders on any directory in your machine. In my case I have created New Folder(2) on desktop and sortedfile folder inside New Folder(2). These paths have been used Mergesort.ipynb

Mergesort: (Mergesort.ipynb)

We are First loading the data set into split function and splitting the data into 12 csv files which are unsorted. Then we are giving these 12 files as input to our merge sort function and the merge sort function will take 2 files each of the unsorted files then perform merge sort on them like this for first iteration 6 Merge sorts are happening between 2 files each and then those files will get merged into 3 files with sorted data and we deleting those 6 files and again running it will run on second iteration there will be 3 files on which we will take 2 files into merge sort and sort them as single file and delete those 2 files. Now we will take the sorted file and merge it with remaining 1 file and perform Merge sort. where it will again compare in same manner it will result in one large sorted file. After all iterations we will get a large sorted file with title in sorted order where alphabets appear first, then small alphabets and special characters as this order was mentioned by the professor. This large file will get divided into 12 csv files with help pf our split function with order mentioned above.

Note : in order to execute the Mergesort.ipynb

- Create two folders on any directory in your computer.
- Give the path of dataset.csv in split function to df and for root path give the file path where you want to store the resulting 12 unsorted csv files.
- In outputFileLambda give the path where you want the sorted files to be stored (there will 12 sorted files ).
- So while executing the all cells in Mergesort.ipynb after changing the path. At the end you will have 12 files created as path mentioned in outputFileLambda.

Linear Search: (Linearsearch.ipynb)

We are first printing the present time when we execute the linearsearch function.We are searching for the title "Sandman: Dream Hunters 30th Anniversary Edition" in all csv files by iterating over each file and searching whether we have the title "Sandman: Dream Hunters 30th Anniversary Edition" in the file or not. If we don't have the tittle we are taking the next csv file and searching for the title. If the title was found we will print the file path where it was found and the row number in that csv file where the record was found and entire data present in that row and we are printing the time when we found the title. The sample image can be seen below,

```
2022-11-02 19:36:10.023627

Located in: C:/Users/mohit/OneDrive/Desktop/New folder(2)/sortedfile/file_sorted_8.csv
Row: 53864
['[38180, 223787]', '77737.0', '[364, 2633]', "A humble young monk and a magical, shape-changing fox find themselves romanti
cally drawn together. As their love blooms, the fox learns of a devilish plot by a group of demons and a Japanese emperor to
steal the monk's life. One of the most popular and critically acclaimed graphic novels of all time, Neil Gaiman's award-winn
ing masterpiece The Sandman set the standard for mature, lyrical fantasy in the modern comics era. Illustrated by an exempla
ry selection of the medium's most gifted artists, the series is a rich blend of modern and ancient mythology in which contem
porary fiction, historical drama, and legend are seamlessly interwoven. In The Dream Hunters, legendary artist P. Craig Russ
ell returns to the world of the Dreaming to adapt Neil Gaiman and artist Yoshitaka Amano's original illustrated novella into
a mesmerizing new comics interpretation. Retelling Gaiman's celebrated story of a humble young monk and the magical fox who
falls in love with him. Russell beautifully brings to life every aspect of the richly evocative world set down in this timel
ess fable of desire, sacrifice, and love that was never meant to be. OVERTURE", '168.0', '259.0', '7.62', '', '30th Annivers
ary ed.', '', '1.0', '9781401294236', '', 'fe69ff7a3776ac320e66bc417db727d3', 'full/b/3/f/b3f35ab174a4486875014b30864dc77811
241515.jpg', 'https://d1w7fb2mkkr3kw.cloudfront.net/assets/images/book/lrg/9781/4012/9781401294236.jpg', '', '', '140129423
5', '9781401294236', 'en', '2019-09-24 00:00:00', '', '4.46', '10168.0', 'Sandman: Dream Hunters 30th Anniversary Edition',
'/Sandman-Dream-Hunters-30th-Anniversary-Edition-Neil-Gaiman/9781401294236', '258.55']

2022-11-02 19:36:22.352764
```

Note: some instructions to execute linear search.ipynb

- After executing Mergesort.ipynb we will have 12 csv files which are sorted. Now in “Data”paste the location where the 12 csv files .

HashIndex: (hashindex.ipynb)

We are first printing the present time when we execute the HashIndex function . Then we have created a dictionary results. We are then iterating over the sorted files. We are taking the title data from the file in which we are in and converting into hash and then we are putting the dictionary with {key, value} in a format where the key is the hash code of title and value is the entire row data. While creating the hash code we are storing the hash code for title: "Sandman: Dream Hunters 30th Anniversary Edition" and after that we are checking if the dictionary result has the desired key that is hash code of title: "Sandman: Dream Hunters 30th Anniversary Edition" if the key was found we are printing the hash code for the title "Sandman: Dream Hunters 30th Anniversary Edition" and its row data and the time when it was found. The sample image can be seen below,

```
2022-11-02 19:36:33.550119
6213193995268118737

Located in: C:/Users/mohit/OneDrive/Desktop/New folder(2)/sortedfile/file_sorted_8.csv
6213193995268118737
['[38180, 223787]', '77737.0', '[364, 2633]', "A humble young monk and a magical, shape-changing fox find themselves romanti
cally drawn together. As their love blooms, the fox learns of a devilish plot by a group of demons and a Japanese emperor to
steal the monk's life. One of the most popular and critically acclaimed graphic novels of all time, Neil Gaiman's award-winn
ing masterpiece The Sandman set the standard for mature, lyrical fantasy in the modern comics era. Illustrated by an exempla
ry selection of the medium's most gifted artists, the series is a rich blend of modern and ancient mythology in which contem
porary fiction, historical drama, and legend are seamlessly interwoven. In The Dream Hunters, legendary artist P. Craig Russ
ell returns to the world of the Dreaming to adapt Neil Gaiman and artist Yoshitaka Amano's original illustrated novella into
a mesmerizing new comics interpretation. Retelling Gaiman's celebrated story of a humble young monk and the magical fox who
falls in love with him. Russell beautifully brings to life every aspect of the richly evocative world set down in this timel
ess fable of desire, sacrifice, and love that was never meant to be. OVERTURE", '168.0', '259.0', '7.62', '', '30th Annivers
ary ed.', '', '1.0', '9781401294236', '', 'fe69ff7a3776ac320e66bc417db727d3', 'full/b/3/f/b3f35ab174a4486875014b30864dc77811
241515.jpg', 'https://d1w7fb2mkkr3kw.cloudfront.net/assets/images/book/lrg/9781/4012/9781401294236.jpg', '', '', '140129423
5', '9781401294236', 'en', '2019-09-24 00:00:00', '', '4.46', '10168.0', 'Sandman: Dream Hunters 30th Anniversary Edition',
'/Sandman-Dream-Hunters-30th-Anniversary-Edition-Neil-Gaiman/9781401294236', '258.55']

2022-11-02 19:36:48.610420
```

Note: some instructions to execute hashindex.ipynb

- After executing Mergesort.ipynb we will have 12 csv files which are sorted. Now in “Data”paste the location where the 12 csv files .

My Laptop specifications:

Processor: Intel core i7-7700hq

Ram size: 16gb ram

Disk storage: 512 gb ssd

Graphic card: Nvidia gtx 1050

Operating System: Windows 10

The entire assignment was done in jupyter notebook.