

## **Programming Assignment 02**

### **Splitting Files**

To split the dataset into numerous files, use the Pandas and NumPy frameworks, then read the dataset.csv file and split it into 100,000-line files, around 12 in total, then save them in sequential order.

### **Merge-sort and linear search**

Merge-sort is a sorting algorithm that employs the divide-and-conquer strategy. Linear search is a sequential search strategy that starts at one end and traverses each list member until the desired section is discovered.

Time taken to find the title: 38.7s

### **Hash index**

A hash index is a data structure used to speed up database queries. I read all of the files and incorporated them into the map to make a hash map. Titles and rows are the keys and values. We can locate the row for the title using a map.

Time taken to find the title: 0.3s

**Machine used:** MSI GF75 i7 10th Generation, 32GB RAM using Windows 11 and run on Visual Studio Code.

**Git link:** <https://git-classes.mst.edu/ap43n/db-assignment-2>