

Jeremy Yang 35769719

Anuj Shah 32046445

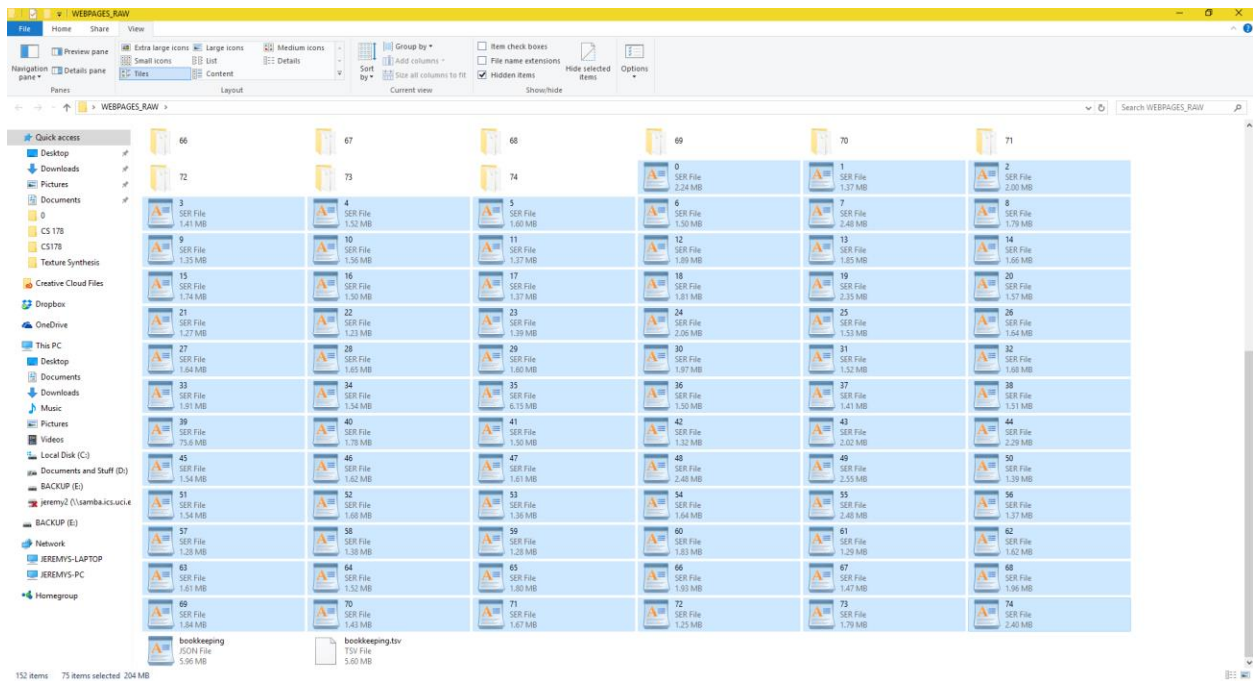
Jack Murray 91682464

## CS 121: Project 3 - Milestone 1

### Flavor: (4) Algorithms and Data Structures Developer

Number of Documents	<b>37497</b>
Number of (Unique) Words	<b>1792732</b>
Total Size of Index	<b>204000 KB</b>

### Additional Notes:



We stored our index in 75 files. Each .ser file corresponds to a folder (i.e. 0.ser represents all the documents inside the 0 directory).

To reduce the time it takes to index the dataset, we used multiple threads. To efficiently store all of the document information, we built and inserted the data into a HashMap.