



CategoryCount.java Counts the number of documents present for each category.

WCountPerCategory.java calculates how many documents contain the term per category.

ChSquarePrepro.java takes the output of WCountPerCategory and sets the term as the key, and a list of categories with the number of documents containing that term as the value, so there is one row per term.

ChiSquare.java read in the output from ChiSquarePrepro and the output from CategoryCount.java to calculate the chi-square value per term per category.

SortingCategory.java just sorts the term per category according to their chisquare value, which i spart of the key. Thanks to a partitioner a separate file for each category is created (so there are 22 reducers used, each containing a sorted list of key value pairs).

finalOutput.java just creates the sorted category and chisquare value list for the final output file.