Word Groups (midterm exam)

Calculate statistics for groups of words which are equal up to permutations of letters. For example, ‘emit’, ‘item’ and ‘time’ are the same words up to a permutation of letters. Determine such groups of words and sum all their counts. Apply stop words filter and convert all words to lowercase letters.

```
stopWords = {"the", "a", "an", "and", "of", "to", "in", "am", "is", "are",
"at", "not"}
wordDelimiters = " \t,;.?!-:@[](){}_*/'"
```

Output record: number of unique words in the group, count of occurrences for the group of words, list of the words in the group in lexicographical order:

groupSize <tab> occurrences <tab> word1 word2 word3 ...

Example: assume ‘emit’ occurred 3 times, 'item' 2 times, 'time' 5 times; 3 + 2 + 5 = 10; group contains 3 words, so for this group result is:

```
3     10    emit item time
```

Find the top 11 groups of words within *Frankenstein; Or, The Modern Prometheus* by Mary Wollstonecraft Shelley (`/user/adampap/Frankenstein`)

The top 4 groups of words within Shakespeare’s books (`/user/adampap/tutorial`) are shown below:

```
4       155     ist its sit tis
4       16      post pots spot stop
4       21      rowse sowre swore worse
4       9       rost rots sort stor
```

The sequence of entries with respect to the second column is not determined.