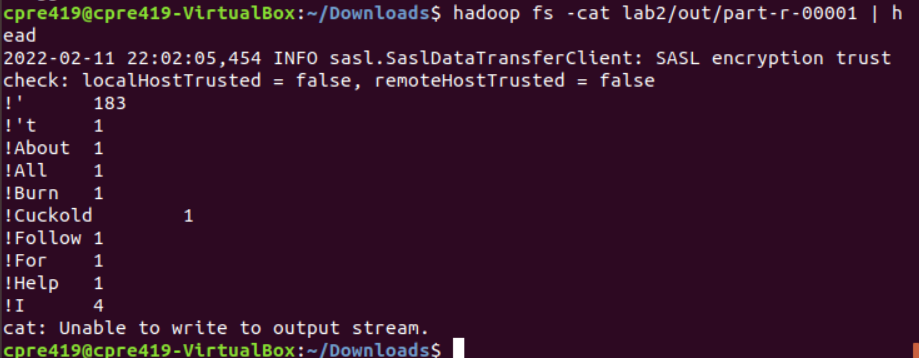
Experiment 1:



First 10 lines in the second output file for wordcount. Could do the same for part-r-00000.

Experiment 2:

This MR program was far more complex than Experiment 1. It required a custom key that stored a bigram and its count. Found the WritableComparable<T> class online and implemented a custom BigramCount key sorted by count. Otherwise, the second map could’ve used count as a key, and have reduce output Bigrams in the format count bigram1, bigram2, bigram3 etc.

In hindsight, the second option would’ve been better. However, I don’t mind having learned some cool Java.

Question Answer: I would write a MR program to preprocess the text file. It would put all sentences on their own respective lines. Then I could write a slightly simplified (knowing that sentences always end at end of line) version of Experiment 2’s code.