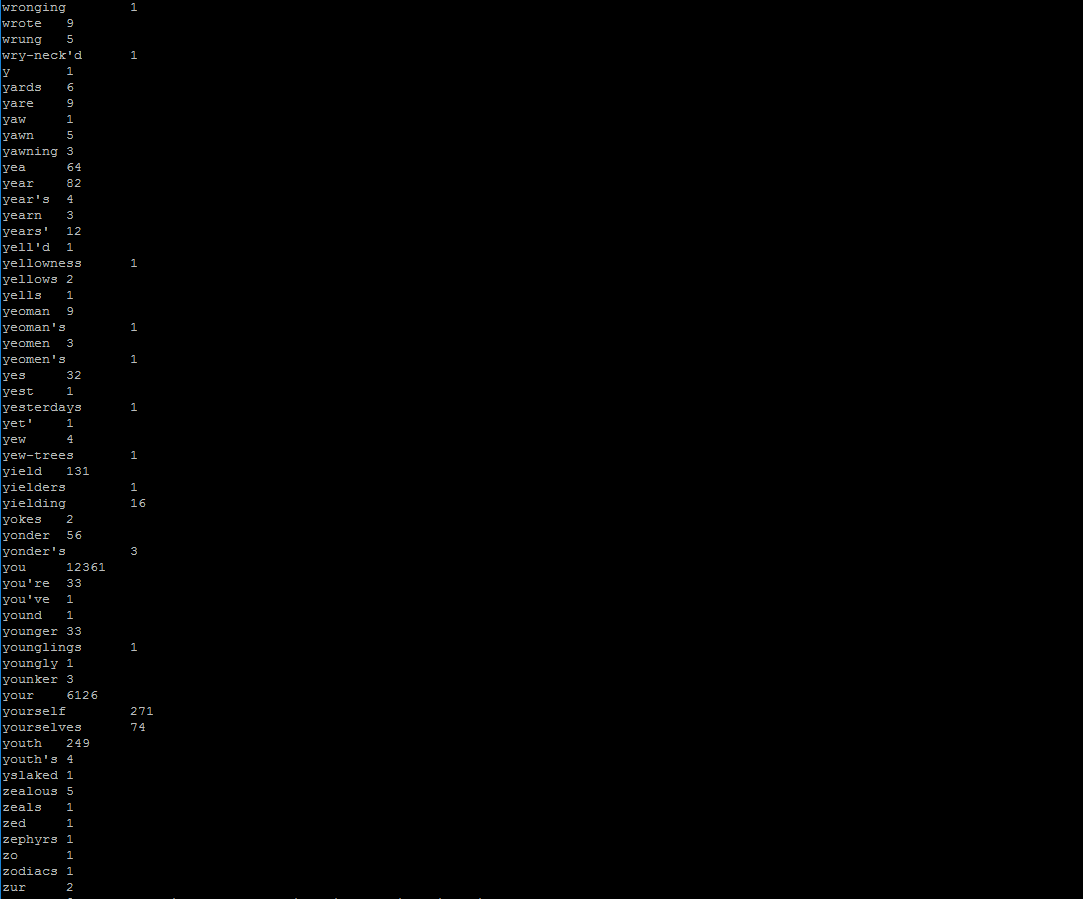
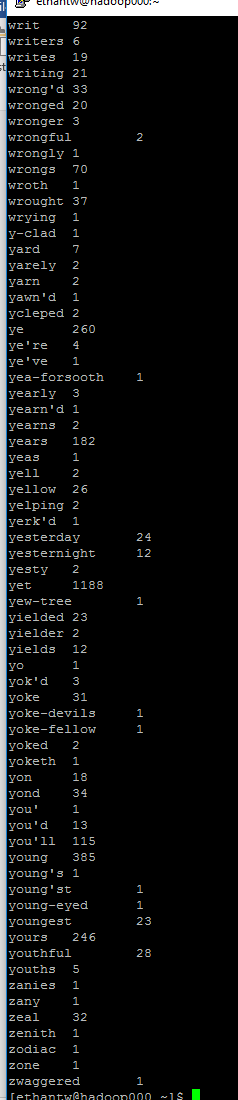
Experiment 1

Part-r-00000

Part-r-00001



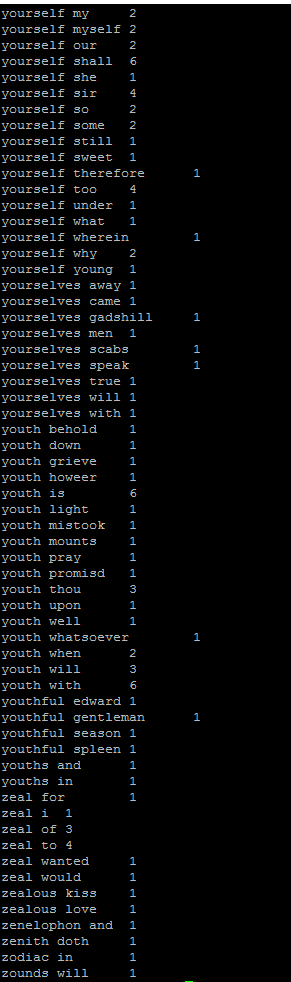
Experiment 2:

Question: Think about how you might be able to get around the fact that bigrams might span lines of input. Briefly describe how you might deal with that situation? (5 points)

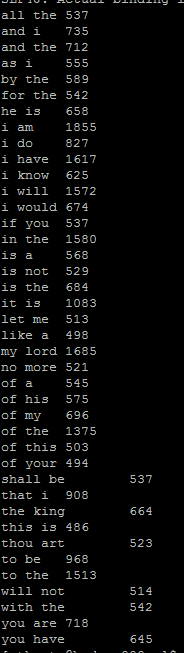
Answer:

One of the ways this could be mitigated would be by having the mapper get passed sentences instead of lines. Or by parsing through the file and storing one sentence per line.

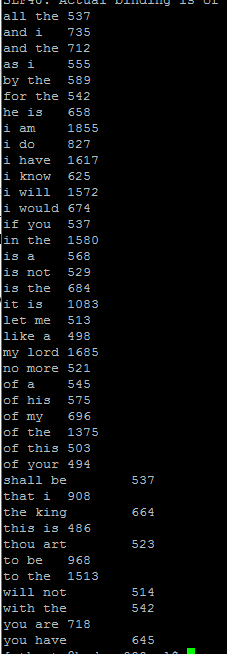
Bigrams:



Top values from Shakespeare:



Top values from Gutenberg:



For some reason. My reducer never gets called or has an issue where the output from the mapper is never removed from the output file. I cannot figure out why this is happening.