Analysis Questions:

1. The average document length is 1215.1629 words per storyID. The smallest storyID : 19406-art53 only has 4 words while the largest storyID: 8951-id_6 has 26139 words in total.
2. Word: ‘the’ appears in most stories. It shows up in ‘966’ stories
   Word: ‘the’ occurs the most number of times. It shows up ‘96151’ times.
3. There are a total of 27217 unique words in total. The words that appear only once are 10056 in total.
   The percent is 37% (unique words that appear only once/ unique words * 100)
   I did not expect this since Heap's law tells us that the rate of distinct words would slowly go down since the words’ data storage becomes larger and larger.
4. The top 100 from the both we get 100 or the top from the both we get 20. We would look at all of the 25 documents and judge them based on the text.