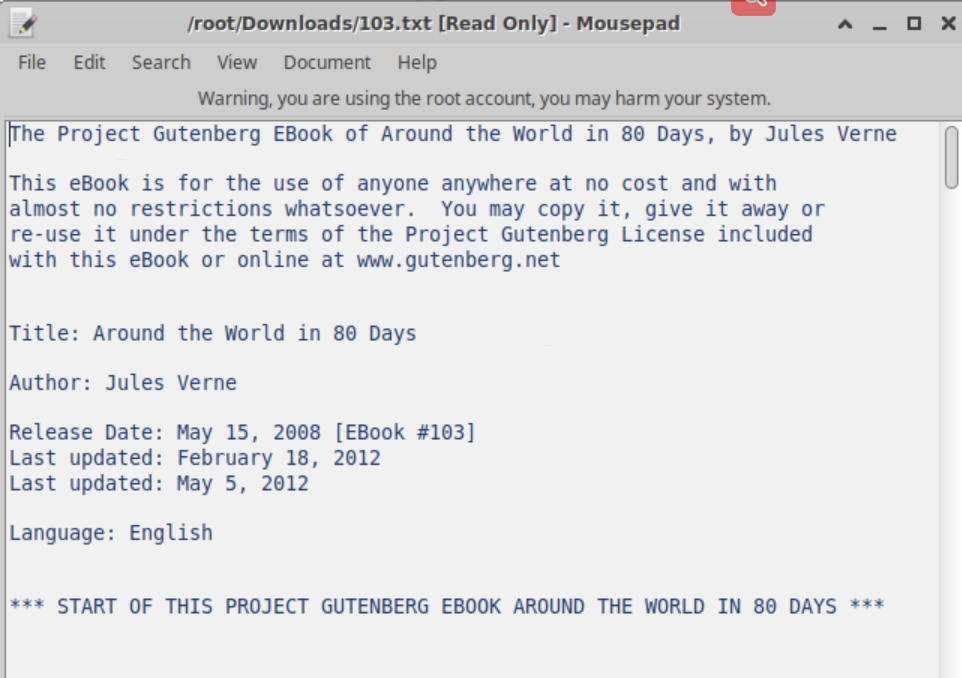
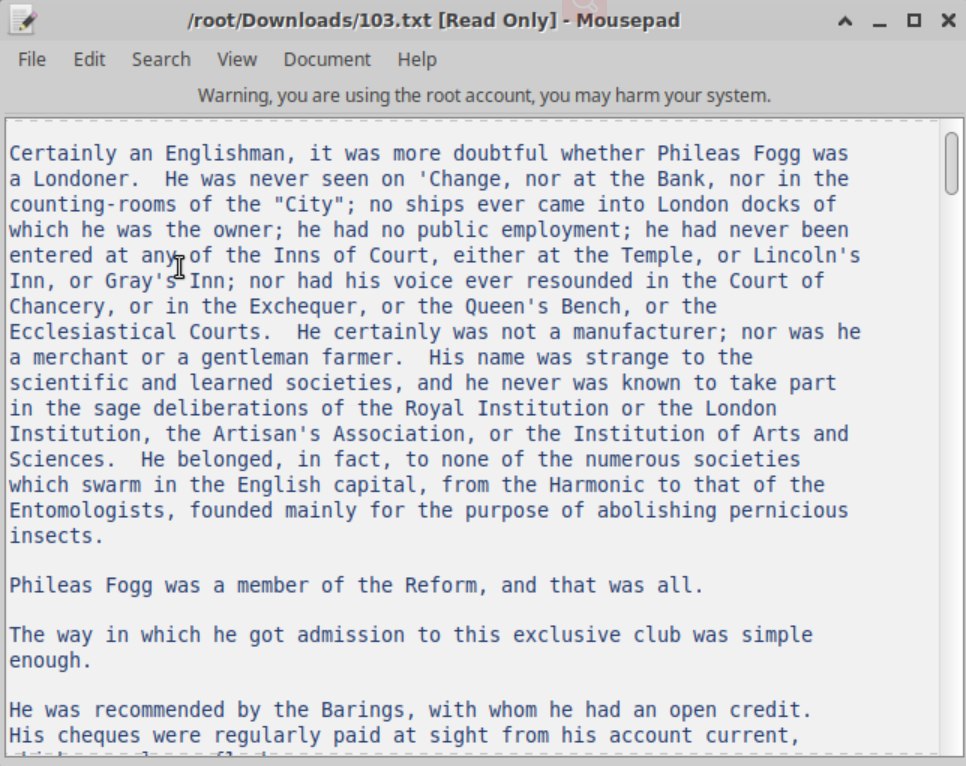
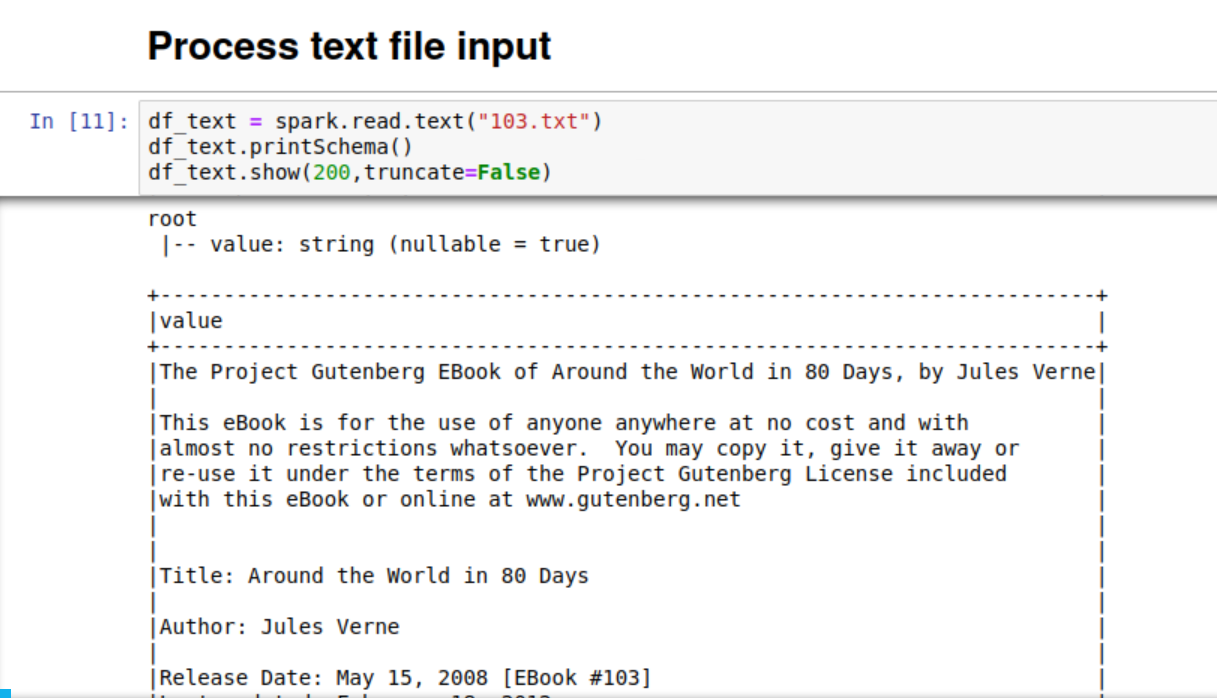
**Intermediate**



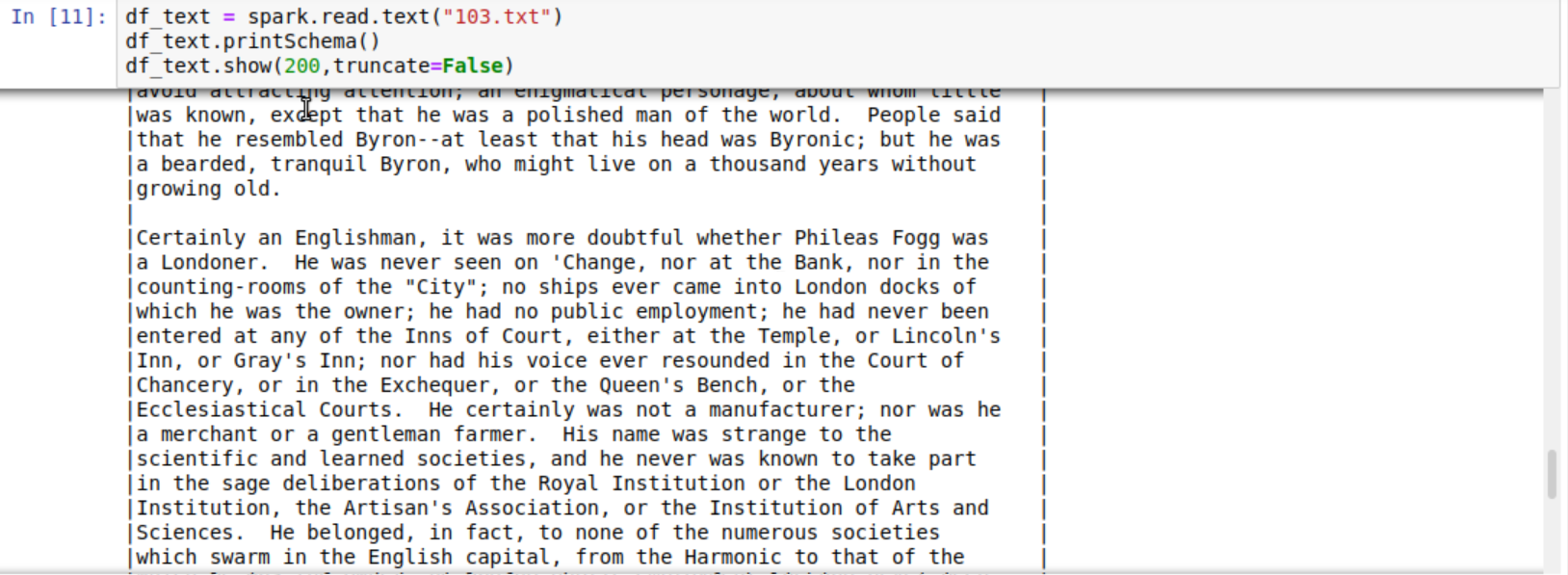
Find an English text book on Gutenberg. (Around the world in 80 days)



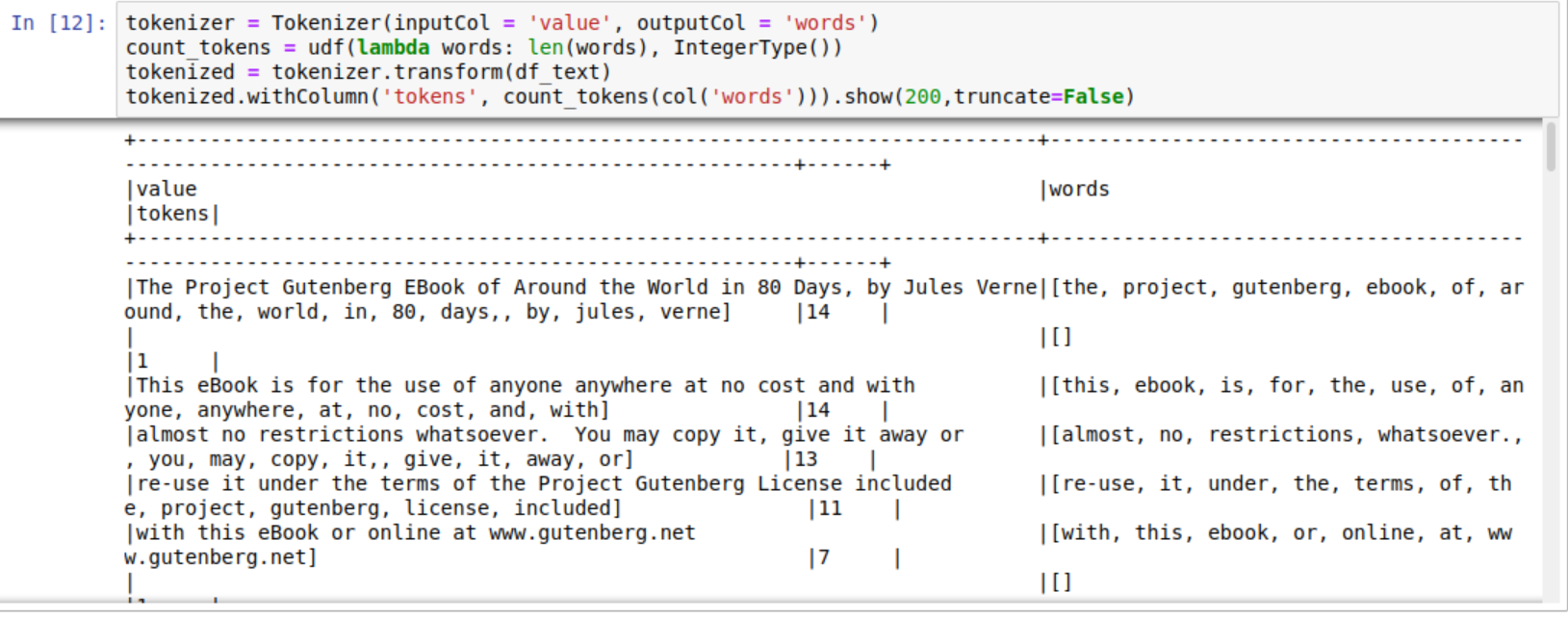
Here is the partial content of the text file



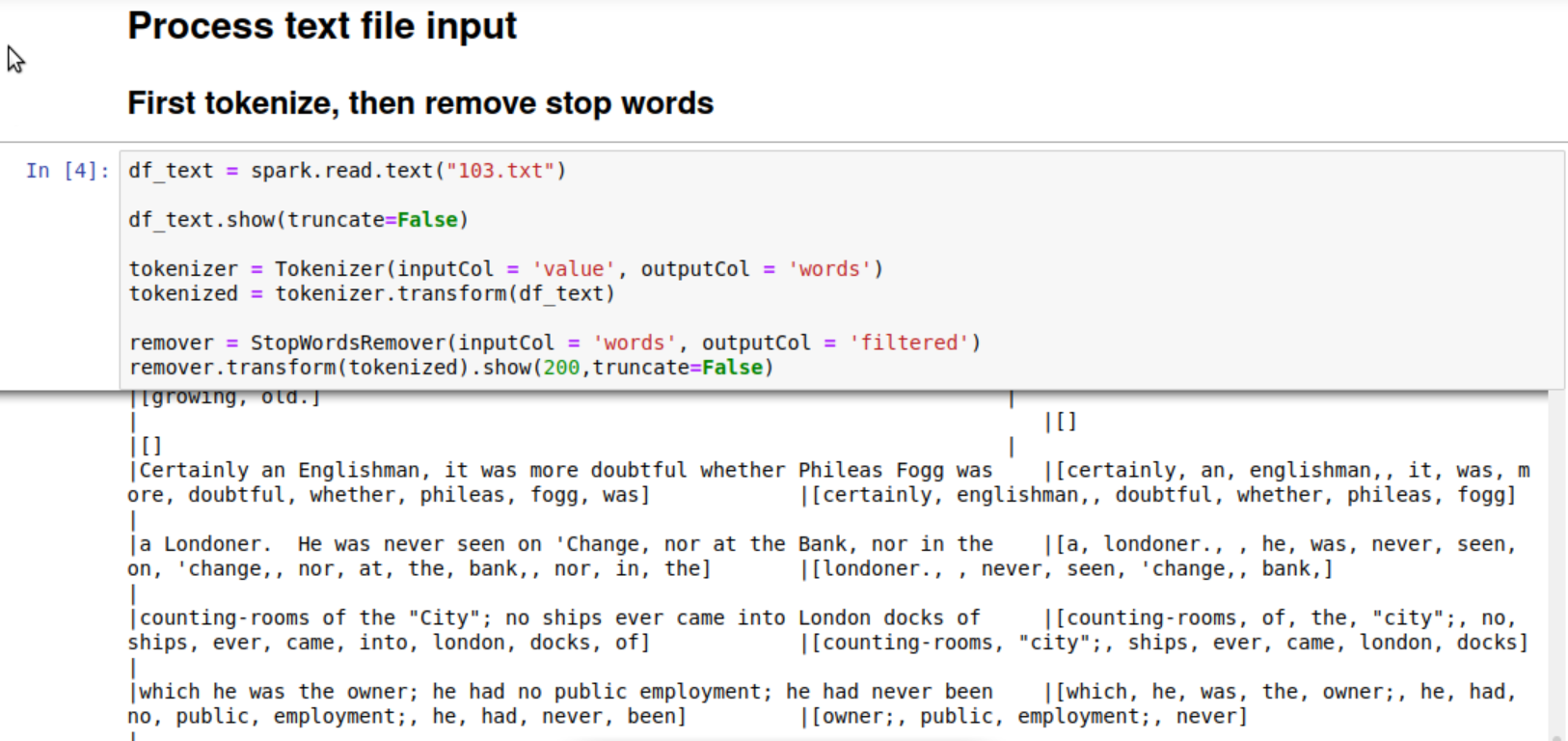
Input such text file as a dataframe in spark



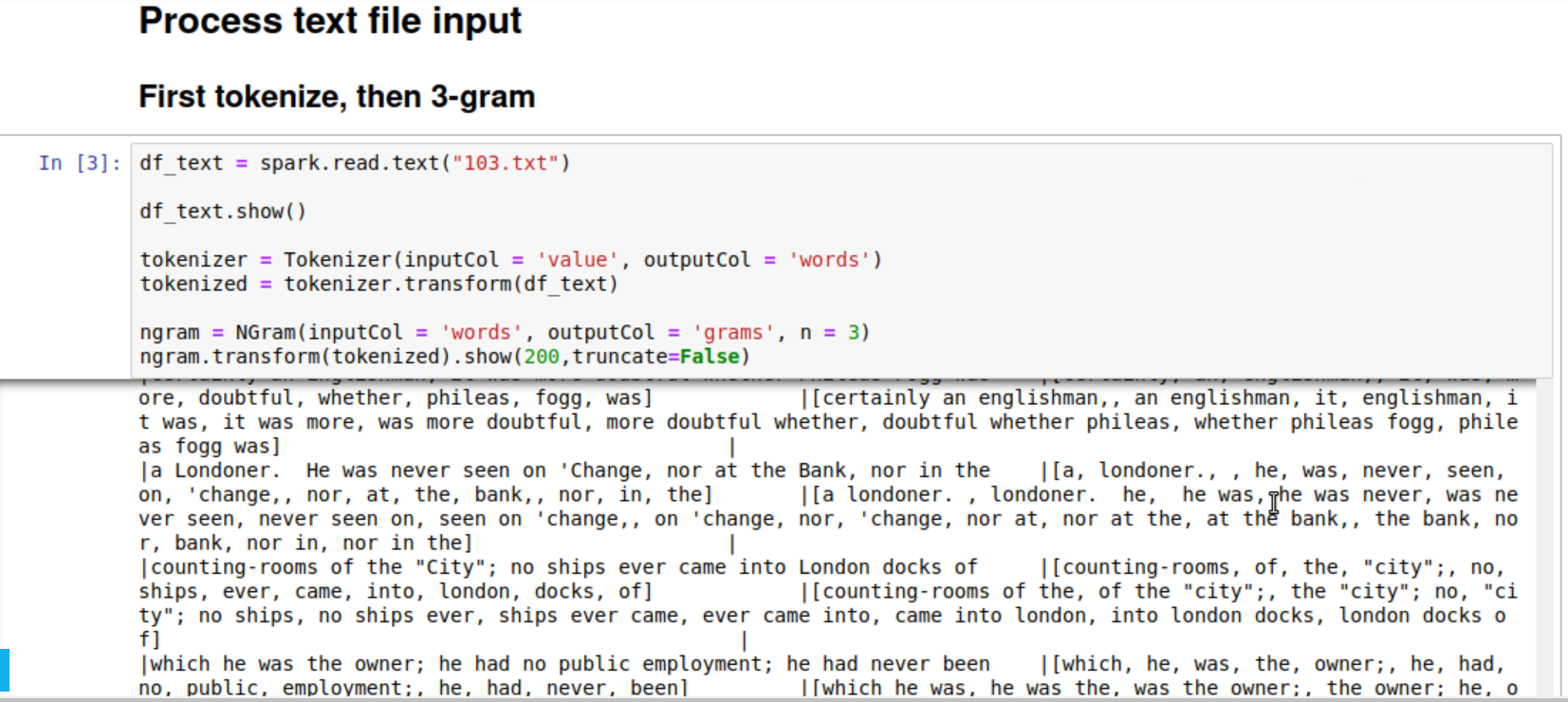
Here is the dataframe content.



First apply tokenization.



Then remove stop words by the built-in function



Get 3-gram of the text file