

Connor Gibbons
CS 446

Project 3 Readme.pdf

1. The average document has 1,215 tokens. The shortest document is 19406-art53 which has 4 tokens, the longest is 8951-id_6 with 26139 tokens.
2. "The" occurs in the most stories, 966. It's also the token with the largest number of occurrences, 96151.
3. There are 27217 different tokens. 10056 of them, or 37%, occur only once.