Eric Mestiza

CSC 161

Assignment 6

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | unknown1.txt &  wss.txt | unknown1.txt &  wbp.txt | unknown2.txt  &  wss.txt | unknown2.txt  &  wbp.txt | unknown3.txt  &  wsce.txt | unknown3.txt  &  bsam.txt | unknown4.txt  &  wsce.txt | unknown4.txt  &  bsam.txt |
| 25 | # of common words: 13  % of common words: 0.52 | # of common words: 12  % of common words: 0.48 | # of common words: 10  % of common words: 0.4 | # of common words: 11  % of common words: 0.44 | # of common words: 16  % of common words: 0.64 | # of common words: 16  % of common words: 0.64 | # of common words: 19  % of common words: 0.76 | # of common words: 16  % of common words: 0.64 |
| 50 | # of common words: 27  % of common words: 0.54 | # of common words: 21  % of common words: 0.42 | # of common words: 18  % of common words: 0.36 | # of common words: 23  % of common words: 0.46 | # of common words: 30  % of common words: 0.6 | # of common words: 38  % of common words: 0.76 | # of common words: 37  % of common words: 0.74 | # of common words: 34  % of common words: 0.68 |
| 100 | # of common words: 44  % of common words: 0.44 | # of common words: 36  % of common words: 0.36 | # of common words: 30  % of common words: 0..3 | # of common words: 43  % of common words: 0.43 | # of common words: 59  % of common words: 0.59 | # of common words: 76  % of common words: 0.76 | # of common words: 78  % of common words: 0.78 | # of common words: 64  % of common words: 0.64 |
| 200 | # of common words: 90  % of common words: 0.45 | # of common words: 64  % of common words: 0.32 | # of common words: N/A  % of common words: N/A | # of common words: N/A  % of common words: N/A | # of common words: 112  % of common words: 0.56 | # of common words: 152  % of common words: 0.76 | # of common words: 143  % of common words: 0.715 | # of common words: 118  % of common words: 0.59 |

N/A=Index out of range error

1.Based on the data unknown1.txt was written by Shakespeare.

2.Based on the data unknown2.txt was written by Blake.

3.Based on the data unknown3.txt was written by Shaw.

4.Based on the data unknown4.txt was written by Shakespeare.

Experimenting with different N values helps to confirm the authors of the unknown texts. This is because as the number of top words being compared increased so did the number of words in common and the percentage of words in common. Ultimately, the number of words in common and the percentage of words in common for one author were higher than the other author and this remained mostly consistent for increasing values of N.