- Method 1 has 78.46153846153847 percent correct
- Method 2 has 78.46153846153847 percent correct
- Method 3 has 83.07692307692308 percent correct

- Method 2 is TF-IDF. Compared to method 1, method 2 squashes the raw counts of terms in each document (raw frequency) by using the $\log_{10}$ of the frequency. Because a word appearing 100 times in a document doesn't make that word 100 times more likely to be relevant to the meaning of the document. In addition, we add 1 to the count because we can't take log of 0. IDF is used to give a higher weight to words that occur only in a few documents. Terms that are limited to a few documents are useful for discriminating those documents from the rest of the collection.

  Method 3 is a TF-IDF variant called maximum tf normalization, which normalize the tf weights of all terms occurring in a document by the maximum tf in that document. The algorithm is as follows, and I set a to 0.4 (generally set):

$$\mathrm{ntf}_{t,d} = a + (1-a)\frac{\mathrm{tf}_{t,d}}{\mathrm{tf}_{\max}(d)},$$

  The main idea of maximum tf normalization is to mitigate the following anomaly: we observe higher term frequencies in longer documents, merely because longer documents tend to repeat the same words over and over again. Since this method makes improvement on the previous TF-IDF, the percent of correctness is higher than method 1&2.