

## Meeting Minute 26/10/2014

1. Read one txt file after another, and store the word frequency into a matrix, which may have large size(column-file, row-word, cell-frequency, each category one matrix)
2. Calculate the information gain(entropy) of the words according to the matrix
3. Creating vectors containing the result which can be obtained from computing the frequency in all other files.
4. Will have several documents for vectors, matrixes

### Elements used

1. Word frequency matrix -> to get information gain matrix
2. Sorting of the information gain, get top frequency words
3. Get the vector with formula