DMAssignment1

PARTH PATEL - 0003574269

February 4, 2016

** Problem 1**.

a :

The most important advantage of tf-idf is that it reduces the length of any big corpus by a substantial amount by removing the frequently occuring stop words like 'the', 'so' etc. The simple tf-idf formula : $x_{ij} = \frac{m_{ij}}{m_i}.log\frac{n}{n_j}$