IDF Weighting: Penalizing Popular Terms

IDF(W)

total number of docs in collection

$$IDF(W) = \log[(M+1)/k]$$

log(M+1)

total number of docs containing W
(Doc Frequency)

1

M

k (doc freq)