

Binary weighting scheme has following advantages:

1. Simplicity
2. Efficiency

Cons:

1. Lower accuracy

Raw count advantage:

1. Better than Binary scheme in accuracy

Disadvantages;

1. Biased towards long docs
2. Sensitive to stop words

Term frequency advantages:

1. Is normalized against length of document => higher efficiency

Cons:

1. Higher computation needed
2. Higher memory needed

Log normalization advantages:

1. Reduces impact of high frequency terms
2. Gives more weight to important words that are rare

Cons:

1. Some words that are important but very common (such as the main subject of the topic) may get misrepresented due to normalization

Double normalization advantages;

1. Reduces impact of high frequency terms
2. Reduces effect of length of document

Cons:

1. May overemphasize the importance of non-important words