

# TKI (Temu Kembali Informasi)

IR

Information Retrieval

**[ PEMBOBOTAN TF.IDF ]**

Dokumen diunduh (download) dari <http://temukembaliinformasi.wordpress.com>

## Menghitung Term Frequency (tf)

Term frequency (tf) merupakan frekuensi kemunculan term (t) pada dokumen (d).

Contoh:

Terdapat kalimat:

Saya sedang belajar menghitung tf.idf. Tf.idf merupakan frekuensi kemunculan term pada dokumen. Langkah awal perhitungan tersebut adalah menghitung tf, kemudian menghitung df dan idf. Langkah terakhir menghitung nilai tf.idf. Mari kita belajar!

Catatan: tiap kalimat dianggap sebagai dokumen.

Tentukan nilai tf!

Jawaban:

Jadi dokumen tiap kalimat ditandai:

Saya sedang belajar menghitung tf.idf. Tf.idf merupakan frekuensi kemunculan term pada dokumen. Langkah awal perhitungan tersebut adalah menghitung tf, kemudian menghitung df dan idf. Langkah terakhir menghitung nilai tf.idf. Mari kita belajar!

Tabel tf:

Term (t)	D1 (dokumen 1)	D2	D3	D4	D5
Akhir	0	0	0	1	0
Awal	0	0	1	0	0
Belajar	1	0	0	0	1
Dokumen	0	1	0	0	0
Frekuensi	0	1	0	0	0
Hitung	1	0	3	1	0
Idf	1	1	1	1	0
Kita	0	0	0	0	1
Langkah	0	0	1	1	0
Muncul	0	1	0	0	0
Saya	1	0	0	0	0
Term	0	1	0	0	0
Tf	1	1	1	1	0

## Menghitung document frequency (df)

Document frequency (df) adalah banyaknya dokumen dimana suatu term (t) muncul.

Contoh: Dari soal yang sama pada menghitung tf, tentukan nilai df!

Jawaban:

Nilai df:

Term (t)	df
Akhir	1
Awal	1
Belajar	2
Dokumen	1
Frekuensi	1
Hitung	3
Idf	4
Kita	1
Langkah	2
Muncul	1
Saya	1
Term	1
Tf	4

Atau:

Term(t)	Akhir	Awal	Belajar	Dokumen	Frekuensi	Hitung	Idf	Kita	Langkah	Muncul	Saya	Term	Tf
Df	1	1	2	1	1	3	4	1	2	1	1	1	4

Menghitung *invers document frequency (idf)*

$$idf = \frac{1}{df}$$

$$\text{Atau: } idf = \log\left(\frac{N}{df}\right)$$

Contoh: Dari soal yang sama pada menghitung df, hitung nilai idf!

Jawaban:

Term (t)	df	idf
Akhir	1	1
Awal	1	1
Belajar	2	$\frac{1}{2}=0.5$
Dokumen	1	1
Frekuensi	1	1
Hitung	3	$\frac{1}{3}=0.3$
Idf	4	$\frac{1}{4}=0.25$
Kita	1	1
Langkah	2	$\frac{1}{2}=0.5$
Muncul	1	1
Saya	1	1
Term	1	1
Tf	4	$\frac{1}{4}=0.25$

Contoh: Dari soal yang sama pada menghitung df, hitung nilai idf! (dengan jumlah dokumen=N)

Jawaban:

Term (t)	df	idf
Akhir	1	$\log(4/1)=0.602$
Awal	1	$\log(4/1)=0.602$
Belajar	2	$\log(4/2)=0.301$
Dokumen	1	$\log(4/1)=0.602$
Frekuensi	1	$\log(4/1)=0.602$
Hitung	3	$\log(4/3)=0.125$
Idf	4	$\log(4/4)=0$
Kita	1	$\log(4/1)=0.602$
Langkah	2	$\log(4/2)=0.301$
Muncul	1	$\log(4/1)=0.602$
Saya	1	$\log(4/1)=0.602$
Term	1	$\log(4/1)=0.602$
Tf	4	$\log(4/4)=0$

Contoh: Dari soal yang sama pada menghitung df, hitung nilai idf! (N=1000)

Jawaban:

Term (t)	df	idf
Akhir	1	$\log(1000/1)=3$
Awal	1	$\log(1000/1)=3$
Belajar	2	$\log(1000/2)=2.70$
Dokumen	1	$\log(1000/1)=0.602$
Frekuensi	1	$\log(1000/1)=0.602$
Hitung	3	$\log(1000/3)=2.52$
Idf	4	$\log(1000/4)=2.40$
Kita	1	$\log(1000/1)=0.602$
Langkah	2	$\log(1000/2)= 2.70$
Muncul	1	$\log(1000/1)=0.602$
Saya	1	$\log(1000/1)=0.602$
Term	1	$\log(1000/1)=0.602$
Tf	4	$\log(1000/4)=2.40$

### Menghitung tf.idf

Hasil kali tf x idf

Contoh: Dari soal yang sama pada menghitung df, hitung nilai tf.idf! (dengan jumlah dokumen=N)

Jawaban:

Term (t)	D1 (dokumen 1)	D2	D3	D4	D5	idf	Tf.idf				
							D1	D2	D3	D4	D5
Akhir	0	0	0	1	0	$\log(4/1)=0.602$	0	0	0	0.602	0
Awal	0	0	1	0	0	$\log(4/1)=0.602$	0	0	0.602	0	0
Belajar	1	0	0	0	1	$\log(4/2)=0.301$	0.301	0	0	0	0.301
Dokumen	0	1	0	0	0	$\log(4/1)=0.602$	0	0.602	0	0	0
Frekuensi	0	1	0	0	0	$\log(4/1)=0.602$	0	0.602	0	0	0
Hitung	1	0	3	1	0	$\log(4/3)=0.125$	0.125	0	0.375	0.125	0
Idf	1	1	1	1	0	$\log(4/4)=0$	0	0	0	0	0
Kita	0	0	0	0	1	$\log(4/1)=0.602$	0	0	0	0	0.602
Langkah	0	0	1	1	0	$\log(4/2)=0.301$	0	0	0.301	0.301	0
Muncul	0	1	0	0	0	$\log(4/1)=0.602$	0	0.602	0	0	0
Saya	1	0	0	0	0	$\log(4/1)=0.602$	0.602	0	0	0	0
Term	0	1	0	0	0	$\log(4/1)=0.602$	0	0.602	0	0	0
Tf	1	1	1	1	0	$\log(4/4)=0$	0	0	0	0	0