

TF - IDF

TF - Term frequency

IDF - Inverse document frequency

sent 1 \Rightarrow good boy
 sent 2 \Rightarrow good girl
 sent 3 \Rightarrow boy girl good

Words frequency

good	3
boy	2
girl	2

Term Frequency Formula

= $\frac{\text{No. of rep of words in sent}}{\text{No. of words in sentence}}$

	TF		
	sent 1	sent 2	sent 3
good	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$
boy	$\frac{1}{2}$	0	$\frac{1}{3}$
girl	0	$\frac{1}{2}$	$\frac{1}{3}$

21

NOVEMBER

THURSDAY

WEEK 47 325-040

S	M	T	W	T	F	S	O
		1	2	3	4	5	C
6	7	8	9	10	11	12	T
13	14	15	16	17	18	19	
20	21	22	23	24	25	26	1
27	28	29	30	31			9

Inverse document frequency

IDF formula.

$$IDF = \log \left(\frac{\text{No. of sentences}}{\text{No. of sent containing words}} \right)$$

Words	IDF
good	$\log(3/3) = 0$
boy	$\log(3/2)$
girl	$\log(3/2)$

TF * IDF

	P_1	P_2	P_3
good		boy	girl
Sent 1	0	$1/2 * \log(3/2)$	0
Sent 2	0	0	$1/2 * \log(3/2)$
Sent 3	0	$1/3 * \log(3/2)$	$(1/2) * \log(3/2)$

boy is given more importance
 girl is " " " "