

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}
d_1	1	1	1	1	1	1	1	0	0	0	0
d_2	1	1	2	0	0	1	1	0	1	0	0
d_3	1	1	1	0	0	0	1	0	0	1	1

normalization: $TF(d, t) = \frac{n(d, t)}{\sum_T n(d, T)} \rightarrow 7$

$TF(1, 1) = \frac{1}{7}$, $TF(1, 2) = \frac{1}{7}$, $TF(1, 3) = \frac{1}{7}$, $TF(1, 4) = \frac{1}{7}$, $TF(1, 5) = \frac{1}{7} = 0.14$

$TF(1, 6) = \frac{1}{7}$, $TF(1, 7) = \frac{1}{7}$

$TF(2, 1) = \frac{1}{7}$, $TF(2, 2) = \frac{1}{7}$, $TF(2, 3) = \frac{2}{7}$, $TF(2, 6) = \frac{1}{7}$, $TF(2, 7) = \frac{1}{7}$

$TF(2, 9) = \frac{1}{7}$

$IDF(t) = \log \frac{1 + |D|}{D_t} = \frac{1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9 \ 10 \ 11}{0.12 \ 0.2 \ 0.2 \ 0.6 \ 0.6 \ 0.3 \ 0.12 \ ?}$

برای جلوگیری از تقسیم بر صفر از فرمول مقابل استفاده می کنیم.

$IDF(t) = \log \frac{1 + |D|}{D_t + 1} = \frac{1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9 \ 10 \ 11}{0.12 \ 0.2 \ 0.2 \ 0.6 \ 0.6 \ 0.3 \ 0.12 \ 0 \ 0.6 \ 0.3 \ 0.3 \ 0.3}$

$TF-IDF(d_t) = TF(d, t) * IDF(t)$

جدول TF-IDF - فضای برداری

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}
$d_1 \leftarrow p$	0	0	0.05	0.04	0.04	0.01	0	0	0	0	0
$d_2 \leftarrow q$	0	0	0.19	0	0	0	0	0	0	0.04	0.04

Cosine Similarity:

$Sim(p, q) = \frac{\vec{p} \cdot \vec{q}}{|\vec{p}| |\vec{q}|} = \frac{(0.05 \times 0.19)}{\sqrt{(0.19)^2 + (0.04)^2 + (0.04)^2} \times \sqrt{(0.05)^2 + (0.04)^2 + (0.04)^2 + (0.01)^2}}$

$= \frac{0.0095}{0.014} = 0.67$