

For pairs $\rightarrow \frac{n(n-1)}{2}$

$\left\{ \begin{array}{l} \text{Ab } O(N^2) \text{ complex} \\ \text{by } O(N) \text{ min} \\ \text{Date: kmay 1st 2019} \\ \text{min-hashing/local} \\ O(N^2) \text{ time} \end{array} \right.$

\Rightarrow Shingling (not for big data)

$D_3 =$ India ~~Not~~ my homeland.

$D_1 =$ Pakistan is my homeland.

$D_2 =$ India is not my homeland.

Main	D_2, D_3		D_1, D_2			D_1, D_3	
	OR	And	D_1	D_3	D_2	And	OR
Pakistan	0	0	1	0	0	0	1
is	1	0	1	0	1	1	1
my	1	1	1	1	1	1	1
homland	1	1	1	1	1	1	1
India	1	1	0	1	0	0	1
not	1	0	0	0	1	0	1
Not	1	0	0	1	0	0	1

Take and of these 2 / Take or of these.

Similarity Jaccard distance (D_1, D_2) = $\frac{|D_1 \cap D_2|}{|D_1 \cup D_2|} = \frac{3}{6} = \frac{1}{2}$

\downarrow and \downarrow or

$\text{Sim}(D_2, D_3) = \frac{3}{6} = \frac{1}{2}$

$\text{Sim}(D_1, D_3) = \frac{2}{6} = \frac{1}{3}$