

BOW:

Words	S_1	S_2	S_3
data	1	1	2
Science	2	1	0
is	1	1	0
one	1	1	0
of	1	1	0
the	1	1	1
most	1	0	0
important	1	0	0
Courses	1	1	0
in	1	0	0
Computer	1	0	0
this	0	0 1	0
best	0	1	0
Scientists	0	0	1
perform	0	0	1
analysis	0	0	1

Bag Of Words

TF

Words	S ₁	S ₂	S ₃
data	1/11	1/8	2/5
science	2/11	1/8	0
is	1/11	1/8	0
one	1/11	1/8	0
of	1/11	1/8	0
the	1/11	1/8	1/5
most	1/11	1/8	0
important	1/11	0	0
courses	1/11	1/8	0
in	1/11	0	0
computer	1/11	0	0
this	0	1/8	0
best	0	1/8	0
scientists	0	0	1/5
perform	0	0	1/5
analysis	0	0	1/5

Term Frequency

idf

Words	S ₁	S ₂	S ₃
data	0	11	11
science	0.18	11	11
is	0.18	11	11
one	0.18	11	11
of	0.18	11	11
the	0	11	11
most	0.48	11	11
important	0.48	11	11
courses	0.18	11	11
in	0.48	11	11
computer	0.48	11	11
this	0.48	0.48	0.48
best		0.48	
Scientists			0.48
perform			0.48
analysis			0.48

Inverse term frequency

TF. Idf

Words

	S ₁	S ₂	S ₃
data	0	0	0
Science	0.033	0.0225	0
is	0.0163	0.0225	0
one	0.0163	0.0225	0
of	0.0163	0.0225	0
the	0	0	0
most	0.0436	0	0
important	0.0436	0	0
courses	0.0163	0.0225	0
in	0.0436	0	0
computer	0.0436	0	0
this	0	0.06	0
best	0	0.06	0
Scientists	0	0	0.096
perform	0	0	0.096
analysis	0	0	0.096

$$\cos(S_1, S_2) = \frac{S_1 \cdot S_2}{|S_1| \times |S_2|}$$

$$S_1 \cdot S_2 = 1 + 2 + 1 + 1 + 1 + 1 + 0 + 0 + 1 + 0 + 0 + 0 + 0 + 0 + 0 + 0$$

$$S_1 \cdot S_2 = 48$$

$$|S_1| = \sqrt{1+4+1+1+1+1+1+1+1+1}$$
$$= \sqrt{14}$$

$$|s_2|^2 = \sqrt{1+1+1+1+1+1+1+1} = \sqrt{9} = 3$$

$$\cos(\sin 52) \approx \frac{48}{354} \approx 0.14$$

$$\cos(s_1, s_2) = \frac{s_1 \cdot s_2}{|s_1| \times |s_2|}$$

$$S_1 S_3 = 2 + 1 = 3$$

~~$18.1 = \sqrt{14}$~~

$$S_3 = \sqrt{8}$$

$$P_S(S_1, S_2) = \frac{3}{\sqrt{14} \times \sqrt{8}} = 0.28$$

$$G_2(S_2, S_3) = \frac{S_2 \cdot S_3}{(S_2) \times (S_3)}$$

$$S_2 \cdot S_1 = 2 + 1 = 3$$

$$|s_2| = \sqrt{9} = 3, |s_3| = \sqrt{8}, \cos(s_2, s_3) = \frac{3}{3\sqrt{8}} = 0.35$$

$$D_{\text{manhattan}}(S_1, S_2) = \sum_{i=1}^{16} |S_1[i] - S_2[i]| \quad \text{for Bow}$$

$$= 1 + 1 + 1 + 1 + 1 + 1 + 1$$

$$= 7$$

$$D_{\text{manhattan}}(S_1, S_3) = \sum_{i=1}^{16} |S_1[i] - S_3[i]|$$

$$= 1 + 2 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1$$

$$= 14$$

$$Q \quad D_{\text{manhattan}}(S_2, S_3) = \sum_{i=1}^{16} |S_2[i] - S_3[i]|$$

$$= 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1$$

$$= 11$$

$$D_{\text{Euclidean}}(S_1, S_2) = \sqrt{D_{\text{manhattan}}(S_1, S_2)}$$

$$= \sqrt{7} = 2.645$$

$$D_{\text{Euclidean}}(S_1, S_3) = \sqrt{D_{\text{man}}(S_1, S_3)}$$

$$= \sqrt{14} = 3.74 \approx 4$$

$$D_{\text{Euc}}(S_2, S_3) = \sqrt{D_{\text{man}}(S_2, S_3)}$$

$$= \sqrt{11} = 3.31$$