SP20-BCS-019 Behzad Khadim IDS-Assign 5

day / date

S1 = " Sunshine state enjoy sunshine ".
S2 = " brown fox jump high, brown fox run ".
S3 = " sunShine state fox run fast ".

Bag of Words

	run	sunShine	state	enjoy	brown	fox	jump	high	fast
S1	0	2	1	1	0	0	0	0	0
S2	1	0	0	0	2	2	1	1	0
S3	1	1	1	0	0	1	0	0	1

Term frequency:-

	run	sunshine	state	enjoy	brown	fox	jump	high	fast
S1	0	2/4	1/4	1/4	0	0	0	0	0
S2	1/7	0	0	0	2/7	2/7	1/7	1/7	0
S3	1/5	1/5	1/5	0	0	1/5	0	0	1/5

	run	sunshine	state	enjoy	brown	fox	jump	high	fast
S1		0.50	0.250	0.250	0	0	0	0	0
S2	0.143	0	0	0	0.286	0.286	0.143	0.143	0
S3	0.200	0.200	0.200	0	0	0.200	0	0	0.200

Finally Taimoor

day / date

IDF:-

terms	idf	
run	0.176	log(3/2)
sunshine	0.176	log(3/2)
state	0.176	log(3/2)
enjoy	0.477	log(3/1)
brown	0.477	log(3/1)
fox	0.176	log(3/2)
jump	0.477	log(3/1)
high	0.477	log(3/1)
fast	0.477	log(3/1)

Tf-idf:-

	S1	S2	S3
run	0 × 0.1766 = 0	1/7 × 0.176 = 0.025	1/5 × 0.176 = 0.035
sunshine	2/4 × 0.176 = 0.088	0 × 0.176 = 0	1/5 × 0.176 = 0.035
state	1/4 × 0.176 = 0.044	0 × 0.176 = 0	1/5 × 0.176 = 0.035
enjoy	1/4 × 0476 = 0.119	0 × 0.476 = 0	0 = 0
brown	0 × 0.477 = 0	2/7 × 0.477 = 0.136	0 = 0
fox	0 × 0.176 = 0	2/7 × 0.176 = 0.050	1/5 × 0.176 = 0.035
jump	0 × 0.477 = 0	1/7 × 0.477 = 0.068	0 × 0.477 = 0
high	0 × 0.477 = 0	1/7 × 0.477 = 0.068	0 = 0
fast	0 × 0.477 = 0	0 × 0.477 = 0	1/5 × 0.477 = 0.095

Finally Taimoor

day / date

Cosine Similarity :-

S1 = " sunshine state enjoy sunshine "

S3 = " sunshine state fox run fast "

vectors :-

S1 = [0, 2, 1, 1, 0, 0, 0, 0, 0]

S3 = [1, 1, 1, 0, 0, 1, 0, 0, 1]

$$\cos(\theta) = \frac{S1 \cdot S2}{|S1| \, |S3|}$$

S1 . S3 = [0×1 + 2×1 + 1×1 + 1×0 + 0 + 0 + 0 + 0 + 0]

= [0 + 2 + 1 + 0 + 0 + 0 + 0 + 0 + 0]

S1 · S3 = 3.

$$|S1| = \sqrt{0 + 2\times2 + 1\times1 + 1\times1 + 0+0+0+0+0} \ldots\ldots$$

= 2.449

$$|S3| = \sqrt{(1\times1) + (1\times1) + (1\times1) + 0 + 0 + (1\times1) + 0 + 0 + (1\times1)}$$

= 2.236

$$\cos(\theta) = \frac{3}{2.449 \times 2.236} = 0.548$$