S_1= How can I be a geologist?

S_2 = What should I do to be a geologist?

Word vector space: [HOW, CAN, I, BE, A, GEOLOGIST, WHAT, SHOULD, DO, TO]

To convert, we put a '1' every where the word occurs, and a '0' if the word does not occur.

Therefore, for **S_1**

HOW	1
CAN	1
I	1
BE	1
A	1
GEOLOGIST	1
WHAT	0
SHOULD	0
DO	0
TO	0

S_1: [1, 1, 1, 1, 1, 1, 0, 0, 0, 0]

For, **S_2**

HOW	0
CAN	0
I	1
BE	1
A	1
GEOLOGIST	1
WHAT	1
SHOULD	1
DO	1
TO	1

S_2: [0, 0, 1, 1, 1, 1, 1, 1, 1, 1]

$\mathbf{S_1} \cdot \mathbf{S_2} = |S_1||S_2|\cos\theta$

So, $\cos\theta$ (which is our cosine similarity) = $(\mathbf{S_1} \cdot \mathbf{S_2})/ (|S_1||S_2|)$

$\mathbf{S_1} \cdot \mathbf{S_2} = 4$
$|S_1| = \sqrt{(1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 0^2 + 0^2 + 0^2 + 0^2)} \cong 2.44949$
$|S_2| = \sqrt{(0^2 + 0^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2)} \cong 2.82843$

$\cos\theta = 4/(2.44949 * 2.82843) \cong 0.5773496$

So, *our cosine similarity between these two sentence vectors is **0.5773496.***
(recall:

- *if the cosine similarity is '0', it means they bear no similarity;*
- *if the cosine similarity is '1', it means they are exactly similar – or the same even)*