

Assume that there is one doc D_{target} : This is the target doc.

And we have three candidate docs as:

D1: This is doc one content.

D2: This is doc two and the length of contents exceeds other two.

D3: Doc three has no overlap with others.

Regardless of punctuations and case, please tell me the order of

1.similarity of these three docs to the target doc, higher similarity first: c

2.cosine similarity of these three docs to the target, higher similarity first: b

3. Euclidean distance of these three docs to the target, lower distance first:

d

4. Manhattan distance of these three docs to the target, lower distance first:

b

a. D3,D2,D1

b. D1,D2,D3

c. D2,D1,D3

d. D1,D3,D2

solution:

$\text{sim}(D1,Dt)=3$

$\text{sim}(D2,Dt)=4$

$\text{sim}(D3,Dt)=1$

$$\cos(D1,Dt)=\frac{3}{\sqrt{5}\times\sqrt{5}} = 0.6$$

$$\cos(D2,Dt)=\frac{4}{\sqrt{5}\times\sqrt{14}} = 0.478$$

$$\cos(D3,Dt)=\frac{1}{\sqrt{5}\times\sqrt{7}} = 0.169$$

euclidean:

$$d(D1,Dt)=\sqrt{4} = 2$$

$$d(D2,Dt)=\sqrt{1 + 6 + 4} = 3.31$$

$$d(D3,Dt)=\sqrt{4 + 6} = 3.16$$

Manhattan:

$$d(D1,Dt)=4$$

$$d(D2,Dt)=8+1=9$$

$$d(D3,Dt)=10$$