

Here is a simplified example of the vector space retrieval model. Consider a very small collection C that consists in the following three documents:

- d1: "new york times"
- d2: "new york post"
- d3: "los angeles times"

Some terms appear in two documents, some appear only in one document. The total number of documents is $N=3$. Therefore, the *idf* values for the terms are:

- angeles $\log_2(3/1)=1.584$
- los $\log_2(3/1)=1.584$
- new $\log_2(3/2)=0.584$
- post $\log_2(3/1)=1.584$
- times $\log_2(3/2)=0.584$
- york $\log_2(3/2)=0.584$

For all the documents, we calculate the *tf* scores for all the terms in C. We assume the words in the vectors are ordered alphabetically.

	angeles	los	new	post	times	york
d1	0	0	1	0	1	1
d2	0	0	1	1	0	1
d3	1	1	0	0	1	0

Now we multiply the *tf* scores by the *idf* values of each term, obtaining the following matrix of documents-by-terms: (All the terms appeared only once in each document in our small collection, so the maximum value for normalization is 1.)

	angeles	los	new	post	times	york
d1	0	0	0.584	0	0.584	0.584
d2	0	0	0.584	1.584	0	0.584
d3	1.584	1.584	0	0	0.584	0

Given the following query: "new new times", we calculate the *tf-idf* vector for the query, and compute the score of each document in C relative to this query, using the cosine similarity measure. When computing the *tf-idf* values for the query terms we divide the



CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
 CALL OR WHATSAPP:689 45 44 70

We calculate the length of each document and of the query:

$$\text{Length of } d1 = \sqrt{0.584^2+0.584^2+0.584^2}=1.011$$

$$\text{Length of } d2 = \sqrt{0.584^2+1.584^2+0.584^2}=1.786$$

$$\text{Length of } d3 = \sqrt{1.584^2+1.584^2+0.584^2}=2.316$$

$$\text{Length of } q = \sqrt{0.584^2+0.292^2}=0.652$$

Then the similarity values are:

$$\text{cosSim}(d1,q) = (0*0+0*0+0.584*0.584+0*0+0.584*0.292+0.584*0) / (1.011*0.652) = 0.776$$

$$\text{cosSim}(d2,q) = (0*0+0*0+0.584*0.584+1.584*0+0*0.292+0.584*0) / (1.786*0.652) = 0.292$$

$$\text{cosSim}(d3,q) = (1.584*0+1.584*0+0*0.584+0*0+0.584*0.292+0*0) / (2.316*0.652) = 0.112$$

According to the similarity values, the final order in which the documents are presented as result to the query will be: d1, d2, d3.

The logo for Cartagena99 features the text 'Cartagena99' in a stylized, teal-colored font. The '99' is significantly larger and more prominent than the 'Cartagena' part. The text is set against a background of light blue and orange geometric shapes, including a large blue triangle and an orange shape that looks like a stylized '9' or a drop.

**CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70**

**ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70**