

ACTO1 – Pregunta 3

(Grupo2_Mañana – 0,6 puntos)

Apellidos y Nombre: Díaz-Alejo León, Stéphane

(IMPORTANTE: todos los cálculos se mostrarán **redondeados a dos decimales**; se deben **justificar** las respuestas)

- 3) Sean una colección de documentos compuesta por los documentos doc1 y doc2 y la consulta:

doc1: Conservación Parque Doñana

doc2: Parque Nacional Garajonay

consulta: Conservación Parque Nacional Garajonay

Todas las palabras de los documentos y la consulta son consideradas como términos para su representación. **Se pide:**

- a) Completar la tabla para un esquema de pesado ltc.ltc (log-pesado, idf y coseno normalizado).

(0,4 puntos)

Term			Consulta				Doc1				Doc2			
	df _t	idf _t	f _{t,d}	tf _{t,d}	w _{t,d} =tf _t idf _d	L-Normaliz	f _{t,d}	tft,d	w _{t,d} =tf _t idf _d	L-Normaliz	f _{t,d}	tf _{t,d}	w _{t,d} =tf _t idf _d	L-Normaliz
Conservación	1	0.3	1	1	0.3	0.58	1	1	0.3	0.71	0	0	0	0
Parque	2	0	1	1	0	0	1	1	0	0	1	1	0	0
Doñana	1	0.3	0	0	0	0	1	1	0.3	0.71	0	0	0	0
Nacional	1	0.3	1	1	0.3	0.58	0	0	0	0	1	1	0.3	0.71
Garajonay	1	0.3	1	1	0.3	0.58	0	0	0	0	1	1	0.3	0.71

- b) Indicar qué documento es más relevante para la consulta en base a la similitud coseno con esquema de pesado ltc.ltc.

(0,2 puntos)

Es más relevante el segundo documento, ya que su similitud coseno con la consulta es 0.82, mayor que la similitud coseno entre el primer documento y la consulta, que es 0.41.

Similitud coseno(Doc1, Consulta) = $0.58 \times 0.71 + 0 \times 0 + 0 \times 0.71 + 0.58 \times 0 + 0.58 \times 0 = 0.41$

Similitud coseno(Doc2, Consulta) = $0.58 \times 0 + 0 \times 0 + 0 \times 0 + 0.58 \times 0.71 + 0.58 \times 0.71 = 0.82$