

RECUPERACIÓN ACT01 – SAR

(17/06/2016)

Apellidos y Nombre:

(IMPORTANTE: todos los cálculos se mostrarán truncados a dos decimales)

1) En una colección de test para una consulta tenemos 7 documentos relevantes . Entre los 10 documentos devueltos sólo 4 son relevantes ocupando las posiciones 2,3,5,8.

Se pide:

- a) Calcula la eficacia del sistema sin tener en cuenta el orden de los documentos en términos de Precisión, Recall, F-medida con $\beta=1$ (No se puntuarán las respuestas que consistan únicamente en el valor resultante)

Precisión= $4/10 = 0,4$

Recall= $4/7 = 0,57$

F-medida= $(2 \times 0,4 \times 0,57) / (0,4 + 0,57) = 0,456 / 0,97 = 0,47$

- b) Completa las Tablas de Precision y Recall Reales (expresando la operación de división realizada y el resultado en decimales, p.e. $2/3 = 0,66$) e Interpoladas.

Tabla Precision&Recall Reales

	1	2	3	4	5	6	7	8	9	10
Relevante		si	si		si			si		
Precisión	0	$0,5 / 2$	$0,66 / 2/3$	$0,5 / 2/4$	$0,6 / 3/5$	$0,5 / 3/6$	$0,42 / 3/7$	$0,5 / 4/8$	$0,44 / 4/9$	$0,4 / 4/10$
Recall	0	$0,14 / 1/7$	$0,28 / 2/7$	$0,28 / 2/7$	$0,42 / 3/7$	$0,42 / 3/7$	$0,42 / 3/7$	$0,57 / 4/7$	$0,57 / 4/7$	$0,57 / 4/7$

Tabla Precision&Recall Interpoladas

Precisión	0,66	0,66	0,66	0,6	0,6	0,44	0	0	0	0	0
Recall	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0

3) Considerando los cuatro documentos siguientes:

Doc1: "No aceptes la admiración de tu perro como una prueba concluyente de que eres maravilloso"

Doc2: "Mi meta en la vida es llegar a ser tan fabuloso como mi perro cree que soy"

Doc3: "Yo no soy maravilloso, maravilloso es mi sabueso"

Se pide:

- a) Completa la tabla tomando como términos "perro" y "maravilloso" usando tf-idf (para el cálculo se toma log-pesado).

Término	Doc1	Doc2	Doc3	df	idf	tf.idf (D1)	Norm(D1)	tf.idf (D2)	Norm(D2)	tf.idf (D3)	Norm(D3)
perro	1	1	0	2	0,17	0,17	0,7	0,17	1	0	0
maravilloso	1	0	2	2	0,17	0,17	0,7	0	0	0,22	1

- b) Calcula la distancia coseno entre Doc1 y Doc2, y entre Doc2 y Doc3, con un esquema de pesado ltc (log-pesado, idf y coseno normalizado).

cos (Doc1 , Doc2) = 0,7

cos (Doc2 , Doc3) = 0

4) Calcula la Distancia de Levenshtein entre las siguientes palabras. Utiliza la cuadrícula para representar los costes acumulados. La cuadrícula tiene un tamaño fijo, que no tiene por qué ajustarse exactamente al espacio que necesitáis utilizar.

D(cala, calanda) = 3

a	4	3	2	1	0	1	2	3
l	3	2	1	0	1	2	3	4
a	2	1	0	1	2	3	4	5
c	1	0	1	2	3	4	5	6
#	0	1	2	3	4	5	6	7
	#	c	a	l	a	n	d	a