

## Recuperación ACTO1 – SAR

(18/06/2018 – 3 puntos)

Apellidos y Nombre: .....

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

- 1) En una colección de test para una consulta tenemos 20 documentos relevantes. Entre los 10 documentos devueltos solamente 4 son relevantes ocupando las posiciones 1, 3, 8, 9.

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). **(0,3 puntos)**

**Precisión=**

**Recall=**

**F-medida=**

- b) Completa las Tablas de Precision y Recall (expresando la operación de división realizada y el resultado en decimales, p.e.  $2/3 = 0,66$ ) e Interpoladas. **(0,6 puntos)**

**Tabla Precision&Recall Reales**

	1	2	3	4	5	6	7	8	9	10
Relevante										
Precisión										
Recall										

**Tabla Precision&Recall Interpoladas**

Precisión											
Recall	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0

- c) Calcula la precisión media. **(0,1 punto)**

2) Considérese la siguiente colección de 3 documentos:

*Doc1*: Cuando hay jamón comen jamón, y si no hay jamón se conforman con pan.

*Doc2*: No les gusta comer.

*Doc3*: Comía jamón como si de pan se tratara.

**Se pide:**

a) Completar la tabla tomando como términos sólo las palabras **jamón, pan** y **comer** (suponiendo que se ha realizado un proceso de lematización sobre los verbos).

**(0,6 puntos)**

Term			Doc1				Doc2				Doc3			
	df <sub>t</sub>	idf <sub>t</sub>	f <sub>t,d</sub>	tf <sub>td</sub>	W <sub>t,d</sub> =tf.xidf	L-Norm	f <sub>t,d</sub>	tf <sub>t,d</sub>	w <sub>t,d</sub> =tf.xidf	L-Norm	f <sub>t,d</sub>	tft,d	w <sub>t,d</sub> =tf.xidf	L-Norm
jamón														
pan														
comer														

b) Considerando la tabla anterior, se pide calcular la similitud coseno entre *Doc1* y *Doc3* con un esquema de pesado lnc (log-pesado, idf y coseno normalizado).

**(0,4 puntos)**

3) Esta pregunta consta de dos apartados:

**(0,6 puntos)**

a) ¿Cómo sería el índice permuterm para la palabra “plaza”?

b) ¿Cómo sería el mecanismo de búsqueda correspondiente a los wildcard queries “pl\*za” y “\*laza” suponiendo que disponemos de este tipo de índice?

4) Se pide completar la inserción en una tabla hash cerrada de tamaño B=11, con función hash  $H(x) = x \text{ MOD } B$ , y con estrategia de redispersión 2ª función hash

$$h_i(x) = (h_{i-1}(x) + k(x)) \text{ MOD } B \text{ siendo } k(x) = (x \text{ MOD } (B-2)) + 1$$

de los siguientes elementos: 51, 14, 3, 7, 18 y 30.

**(0,4 puntos)**

## Soluciones:

### 1.a)

Precisión=  $4/10=0,4$

Recall=  $4/20=0,2$

F-medida= 0,26

### 1.b)

Tabla Precision&Recall Reales

	1	2	3	4	5	6	7	8	9	10
Relevante	Y	N	Y	N	N	N	N	Y	Y	N
Precisión	1	0,5	0,66	0,5	0,4	0,33	0,28	0,37	0,44	0,4
Recall	0,05	0,05	0,1	0,1	0,1	0,1	0,1	0,15	0,2	0,2

Tabla Precision&Recall Interpoladas

Precisión	1	0,66	0,44	0	0	0	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

1.c)  $P_m=(1+0,66+0,37+0,44)/20=0,12$

### 2.a)

Term			Doc1				Doc2				Doc3			
	df <sub>t</sub>	idf <sub>t</sub>	f <sub>t,d</sub>	tf <sub>t,d</sub>	W <sub>t,d</sub> =tf <sub>t,d</sub> idf <sub>t</sub>	L-Norm	f <sub>t,d</sub>	tf <sub>t,d</sub>	w <sub>t,d</sub> =tf <sub>t,d</sub> idf <sub>t</sub>	L-Norm	f <sub>t,d</sub>	tf <sub>t,d</sub>	w <sub>t,d</sub> =tf <sub>t,d</sub> idf <sub>t</sub>	L-Norm
jamón	2	0,17	3	1,47	0,2499	0,82	0	0	0	0,00	1	1	0,17	0,70
pan	2	0,17	1	1	0,17	0,56	0	0	0	0,00	1	1	0,17	0,70
comer	3	0	1	1	0	0,00	1	1	0	0,00	1	1	0	0,00

2.b)  $\cos(\text{Doc1}, \text{Doc3}) = (0,82 \times 0,7) + (0,65 \times 0,7) + (0 \times 0) = 0,96$

3.a) El índice permuterm para el término “plaza” se construiría con las diferentes rotaciones del término:

plaza\$

laza\$p

aza\$pl

za\$pla

a\$plaz

\$plaza

### 3.b)

La búsqueda que se realiza para “pl\*za” es “za\$pl\*” y para “\*laza” es “laza\$\*”.

4)

x	H(x)	$h_1(x)$	$h_2(x)$
51	7		
14	3		
3	3	$(3+3+1) \text{MOD } B=7$	$(7+3+1) \text{MOD } B=0$
7	7	$(7+7+1) \text{MOD } B=4$	
18	7	$(7+8+1) \text{MOD } B=5$	
30	8		

La tabla queda como sigue:

0	1	2	3	4	5	6	7	8	9	10
3			14	7	18		51	30		