

# Acto1 – Preguntas 1 y 2

(Grupo2\_Mañana – 1,4 puntos)

Apellidos y Nombre: Díaz-Alejo León, Stéphane

**IMPORTANTE:** todos los cálculos se mostrarán redondeados a dos decimales

- 1) Sea una colección de documentos con 100 documentos, identificados con los números de 1 al 100. Sabemos que los documentos relevantes para una determinada consulta son:

[3,5,18,22,35,40,41,63,80,89].

Un sistema de recuperación de información devuelve el siguiente resultado para la consulta:

S1= [40, 12, 27, 5, 41, 20, 99, 80, 8, 17, 35, 19]

Se pide:

- a) Calcular la eficacia (Precision, Recall y la F-medida con  $\beta=1$ ) para la consulta. **(0,3 puntos)**

Consulta	Precision	Recall	F-1
S1	$5/12=0.42$	$5/10=0.5$	$(2 \times 0.42 \times 0.5) / (0.42 + 0.5) = 0.46$

- b) Completar las Tablas de Precision y Recall (expresando la operación de división realizada y el resultado redondeado en dos decimales, p.e.  $2/3 = 0,67$ ) e Interpoladas. **(0,6 puntos)**

**Tabla Precision&Recall Reales**

S1	1	2	3	4	5	6	7	8	9	10	11	12
Relevante	Yes	No	No	Yes	Yes	No	No	Yes	No	No	Yes	No
Precision	$1/1=1$	$1/2=0.5$	$1/3=0.33$	$2/4=0.5$	$3/5=0.6$	$3/6=0.5$	$3/7=0.43$	$4/8=0.5$	$4/9=0.4$	$4/10=0.4$	$5/11=0.45$	$5/12=0.42$
Recall	$1/10=0.1$	$1/10=0.1$	$1/10=0.1$	$2/10=0.2$	$3/10=0.3$	$3/10=0.3$	$3/10=0.3$	$4/10=0.4$	$4/10=0.4$	$4/10=0.4$	$5/10=0.5$	$5/10=0.5$

**Tabla Precision&Recall Interpoladas**

Precision S1	1	1	0.6	0.6	0.5	0.45	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

- c) Calcular la R-precision del sistema. **(0,1 puntos)**

R-precision =  $4/10=0.4$

2) Esta pregunta consta de dos apartados:

(0,4 puntos)

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

**gorrión\$, \$gorrión, n\$gorrió, ón\$gorri, ión\$gorr, rión\$gor, rrión\$go, orrión\$g**

Los índices permuterm se crean añadiendo el símbolo “\$” al final de la palabra y rotando una posición todas las letras hasta generar todas las combinaciones.

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

“go\*ión” -> **ión\$go\***

“\*orrión” -> **orrión\$\***

Esto se resuelve añadiendo el símbolo “\$” al final de la wildcard query y rotando las letras hasta lograr posicionar el símbolo “\*” a la derecha.