

Trabajo Práctico Evaluación de la Recuperación

Bibliografía

[MIR] Capítulo 3, [TOL] Capítulo 2, [MAN] Capítulo 8

- 1) Se requiere evaluar la performance en la recuperación de un sistema. Para una consulta q_1 , dicho sistema entregó la siguiente salida.

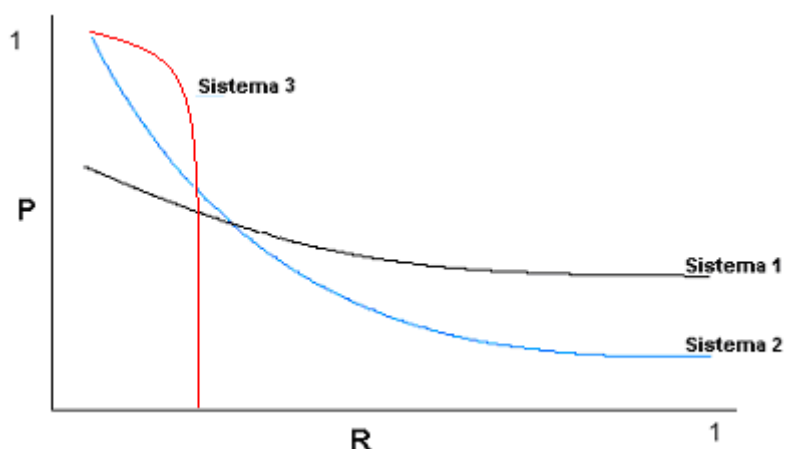
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
R	N	N	R	R	N	N	N	N	R	N	N	N	R	N

Los documentos identificados como R son los relevantes, mientras que las N's corresponden a documentos no relevantes a q_1 . Suponga – además – que existen en el corpus otros 6 documentos relevantes a q_1 que el sistema no recuperó. A partir de esta salida calcule las siguientes medidas:

- a) Recall y Precision para cada posición j
- b) Recall y Precision promedio
- c) Precisión al 50% de Recall
- d) Precisión interpolada al 50% de Recall
- e) Precisión-R

Finalmente, realice las gráficas interpolada y sin interpolar. Luego, interprete brevemente los resultados y brinde una explicación.

- 2) Brinde dos ejemplos en los cuales una persona requiera de un sistema de recuperación de información la mayor precisión y dos en los cuales requiera máxima exhaustividad.
- 3) A partir de la siguiente gráfica de R y P brinde un breve explicación del comportamiento de cada sistema. ¿Cuál es mejor? Proponga y justifique – para cada caso – aplicaciones ejemplo donde sería adecuado el comportamiento mostrado del sistema.



- 4) Utilizando la colección de prueba CISI (<http://www.ri.unlu.edu.ar/recursos/>) y el software Lemur Toolkit (<http://www.lemurproject.org/>) se debe realizar la evaluación del sistema. Para ello, es necesario construir un índice con los documentos de la colección y luego ejecutar



las consultas. Los resultados deben ser comparados contra los juicios de relevancia de la colección utilizando el software trec_eval (http://trec.nist.gov/trec_eval/). Realizar el análisis y escribir un reporte indicando los resultados obtenidos, junto con la gráfica de R–P en los 11 puntos standard.

- 5) Dadas las salidas de los siguientes sistemas evalúe la performance de cada uno utilizando las medidas F y E. Suponga que existen 10 documentos relevantes para el query.

S1	S2	S3	S4	S5	S6
R	N	R	R	N	N
R	R	N	R	N	R
N	N	R	R	N	N
N	R	N	N	N	N
N	N	N	N	R	R
R	R	N	N	R	R
N	R	R	N	R	N
N	N	R	N	R	R
N	N	N	N	R	N
R	N	R	N	R	N

¿Cómo se modifica su evaluación si pondera en E la Precisión al doble de la Exhaustividad? ¿Y si es al triple? Realice – en cada caso – la gráfica correspondientes.

- 6) Se presentan a continuación dos narrativas que corresponden a necesidades de información de un usuario cualquiera. A partir de éstas, debe crear – por cada una – dos consultas (queries) diferentes. Para crear los queries puede incluir términos que no se encuentren en las descripciones y – eventualmente – escribir las consultas en inglés.

N1: *Se necesitan datos acerca de casos de personas atacadas por diferentes especies de tiburones (específicamente).*

N2: *¿Qué atletas obtuvieron records de velocidad en natación en cualquier estilo? Se necesitan documentos que mencionen al atleta, el estilo y su record.*

Luego ejecute las consultas en el metabuscador Dogpile (<http://www.dogpile.com>) y analice los 20 primeros resultados de cada corrida, juzgando – a su criterio – si el documento es relevante a la necesidad de información o no. Arme una tabla con las siguientes columnas: Q#, D#, JR

Donde Q# es el número de consulta (1..4), D# es el número de documento (1..20) y JR es su juicio de relevancia (R – Relevante, N – No relevante).

Finalmente, ejecute nuevamente las consultas en los buscadores Google y Yahoo y evalúe las 50 primeras respuestas contra sus juicios de relevancia. Calcule la Precisión cada 10 documentos y en los 11 niveles standard de Recall. ¿Qué sistema es mejor? ¿Puede justificar su respuesta? ¿Qué debilidades presenta su evaluación?

7) A continuación se presentan las salidas de tres sistemas de recuperación de información para 3 consultas cualquiera y los juicios de relevancia creados por asesores humanos.

Query 1			Query 2			Query 3		
SRI A	SRI B	SRI C	SRI A	SRI B	SRI C	SRI A	SRI B	SRI C
5	7	11	12	16	7	1	5	23
2	16	2	14	23	23	18	24	9
1	9	10	8	3	3	11	13	25
15	18	21	23	13	13	16	19	18
9	4	1	9	4	4	19	17	1
19	17	5	5	14	8	10	12	2
3	8	22	20	17	17	23	22	11
6	5	9	21	11	25	21	4	10
18	11	20	4	7	16	3	1	19
4	15	3	19	21	21	6	8	16
12	12	23	3	18	18	22	11	12
8	10	4	10	10	22	12	3	22
11	6	19	7	19	19	17	16	15
17	2	6	18	22	10	2	23	8
7	19	15	11	20	20	13	25	24
20	20	25	17	1	5	24	18	14
10	21	18	24	2	2	25	2	7
21	25	16	13	12	12	14	20	13
16	22	7	2	15	15	5	14	17
23	1	8	16	5	1	4	10	20
13	23	17	6	8	14	7	6	21
22	13	12	22	6	24	9	21	4
24	24	13	15	9	9	15	15	5
25	3	14	25	25	11	20	7	6
14	14	24	1	24	6	8	9	3

Juicios de Relevancia

D	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Q1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Q2	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1
Q3	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	1	0

Para cada sistema calcule:

- La precisión media
- La precisión media a intervalos de Recall de 20%.
- P@5, P@10, P@20

Exponga un escenario posible y medidas complementarias para decidir qué sistema utilizar.