

Tema 4: Recuperación de información

4.1. Evaluación de la RI

Juan Manuel Fernández Luna

Dpto. Ciencias de la Computación e Inteligencia Artificial jmfluna@decsai.ugr.es

Determinar la calidad de un S.R.I.

En función de:

- Espacio.
- Tiempo (eficiencia).
- Facilidad de uso (usabilidad).
- Calidad de la recuperación (eficacia).

Espacio: tamaño de los índices.

Relación entre el tamaño de la colección original y los ficheros creados tras la indexación.

Tiempo:

de indexación (procesar la colección) y de recuperación (procesar la consulta).

Todos los criterios anteriores son mesurables, pero...

¿Qué ocurre con la satisfacción del usuario?

¿Qué es?

Velocidad de respuesta, tamaño de índices, son factores de esa satisfacción.

De nada serviría una respuesta totalmente inútil.

Necesitamos una forma de medir la satisfacción.

¿Quién es el usuario al que intentamos satisfacer? Depende del entorno.

Motor de búsqueda web: El usuario encuentra lo que quiere y vuelve a usar el buscador.

Puede medir la tasa de retornos.

Sitio de comercio electrónico: El usuario encuentra lo que quiere y compra.

¿Satisfacción del usuario final o del sitio? Tiempo de compra o fracción de usuarios que compran.

Evaluación de la RI

5

Organización: "Productividad del usuario"

¿Cuánto tiempo ahorran mis empleados cuando buscan información?

Muchas son difíciles de medir...

El más común: la relevancia de los resultados de la búsqueda.

¿Cómo?

3 elementos:

- Colección de prueba.
- Colección de consultas.
- Una asignación de relevancia o irrelevancia para cada para documento-consulta.

Eficacia en la recuperación:

Medidas:

Precisión: proporción de documentos recuperados que son relevantes.

Recall (exahustividad): proporción de documentos relevantes recuperados.

Fallout: proporción de documentos no relevantes recuperados.

Medidas de eficacia

Exhaustividad (Recall)

 mide en qué medida un sistema es capaz de detectar todos los documentos relevantes existentes en la colección en relación a una consulta.

Exhaustividad =
$$\frac{n^{\circ} \text{ de docs relevantes recuperados}}{n^{\circ} \text{ de docs relevantes en la colección}}$$

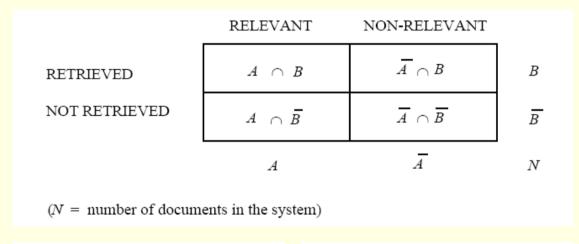
Medidas de eficacia

Precisión.

 mide en qué medida un sistema es capaz de detectar todos los documentos irrelevantes existentes en la colección en relación a una consulta, impidiendo su presentación innecesaria al usuario.

$$Precisi\'on = \frac{n^{\circ} \ de \ docs \ recuperados \ relevantes}{n^{\circ} \ de \ docs \ recuperados}$$

Eficacia en la recuperación:



PRECISION =
$$\frac{|A \cap B|}{|B|}$$
 RECALL = $\frac{|A \cap B|}{|A|}$

FALLOUT =
$$\frac{|\overline{A} \cap B|}{|\overline{A}|}$$

Un sistema de RI, ante una consulta del usuario, contiene en su colección un total de 10 documentos relevantes. El sistema muestra al usuario 4 documentos relevantes y 12 documentos irrelevantes.

¿Exhaustividad y la precisión del sistema?

Exhaustividad =
$$\frac{4}{10}$$
 = 0,4

$$Precisi\'on = \frac{4}{16} = 0,25$$

Recall alto y precisión baja → recuperar todos los documentos.

Eficacia en la recuperación:

Un buen S.R.I. debería

- presentar un recall alto (recuperaría el mayor número posible de documentos relevantes),
- y tener una precisión también alta (recuperar muy pocos no relevantes).

- Existe una relación inversa entre precisión y exhaustividad.
 - Si deseamos aumentar la exhaustividad basta con recuperar más documentos, pero introduciremos documentos no relevantes, disminuyendo así la precisión.
 - Idealmente, es 1 si devolvemos la colección completa
 - Al contrario, si deseamos aumentar la precisión no podemos recuperar muchos documentos, pues tratamos de ajustarnos exclusivamente a los relevantes, pero a cambio dejaremos sin recuperar algún documento relevante.

Algunos inconvenientes:

- Se debería hacer la media a lo largo del conjunto de consultas.
- Se necesita un conjunto de juicios de relevancia realizados por humanos (la gente no suele realizarlos muy bien).
- Los juicios deben ser binarios.
- Dependientes del corpus.

Medida combinada que integra R y P: F1 (media armónica ponderada):

$$F = \frac{1}{\alpha \frac{1}{P} + (1 - \alpha) \frac{1}{R}} = \frac{(\beta^2 + 1)PR}{\beta^2 P + R}$$

Normalmente, F_1 : $\beta = 1$ and $\alpha = \frac{1}{2}$

(C.J. van Rijsbergen, Information Retrieval)

Para poder aplicar los conceptos de exhaustividad y precisión en estos sistemas, se calculan los valores de la exhaustividad y precisión para cada posición de la salida.

1 2 3 4 5 6 7 8

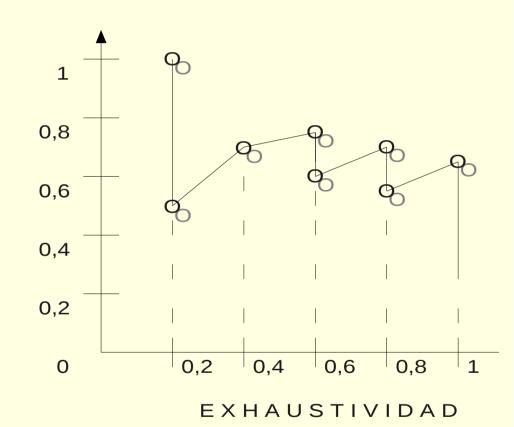
S1: r nr r nr r nr r

	Posic 1	Posic 2	Posic 3	Posic 4	Posic 5	Posic 6	Posic 7	Posic 8
Exhaust	0,2	0,2	0,4	0,6	0,6	0,8	0,8	1
Precis	1	0,5	0,67	0,75	0,6	0,67	0,57	0,63

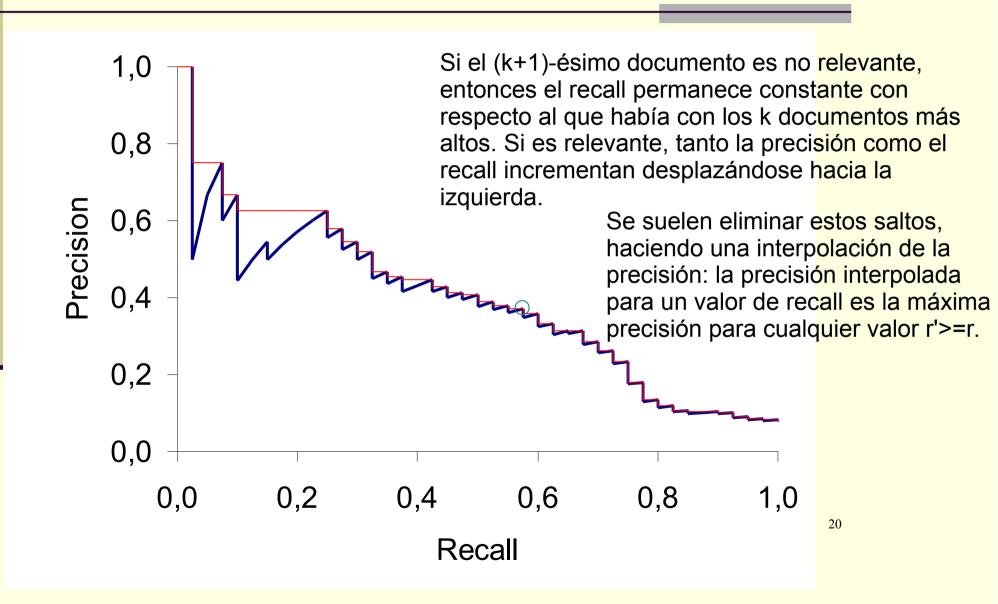
Curvas Recall/Precision

Consiste en representar gráficamente los datos anteriores.

- eje de abscisas (X) la exhaustividad.
- eje de ordenadas (Y) la precisión.
 - se dibuja en el plano el punto (exh,pr).



- Las curvas precisión-exhaustividad poseen una forma típica en diente de sierra.
- Esto es debido a que un sistema de RI recupera habitualmente de manera alternativa documentos relevantes e irrelevantes:
 - Sig doc es Relevante: fragmentos de líneas ascendentes hacia la derecha
 - Sig doc es No Relevante: fragmento de líneas verticales descendentes



Pero...

Un gráfico P-R para una sóla consulta no es muy informativo.

Se debe obtener el rendimiento medio para un conjunto de consultas.

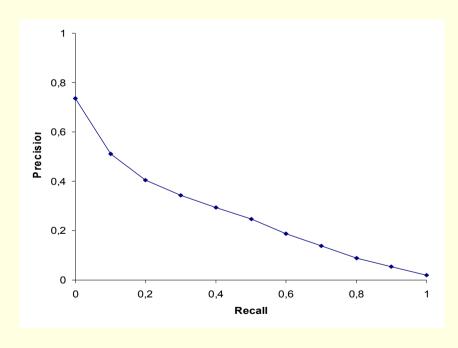
Cuestión técnica:

Los cálculos P-R disponen ciertos puntos en el gráfico.

¿Cómo se determina un valor entre puntos?

Solución: Interpolar los valores de precisión para 11 valores fijos de recall:

0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 y 1.0.



Mejor medidas resumen:

Precisión a un nivel de recuperación fijo (P@x).

- Más apropiado para la búsqueda web: los mejores resultados en la primera y segunda página.
- Valor arbitrario de K.

Precisión media interpolada de 11 puntos de recall.

Evalúa el rendimiento en niveles de recall.

Colecciones de prueba:

24

- Colección de documentos.
- Conjunto de consultas.
- Para cada consulta, el conjunto de documentos estándar.

Ejemplos: TREC, INEX, CACM, CISI, MEDLARS,...

Collection	NDocs	NQrys	Size (MB)	Term/Doc	Q-D RelAss
ADI	82	35			
AIT	2109	14	2	400	>10,000
CACM	3204	64	2	24.5	
CISI	1460	112	2	46.5	
Cranfield	1400	225	2	53.1	
LISA	5872	35	3		
Medline	1033	30	1		
NPL	11,429	93	3		
OSHMED	34,8566	106	400	250	16,140
Reuters	21,578	672	28	131	
TREC	740,000	200	2000	89-3543	» 100,000