

## Parte II

---

### Evaluación de IR

## Evaluación de la recuperación

---

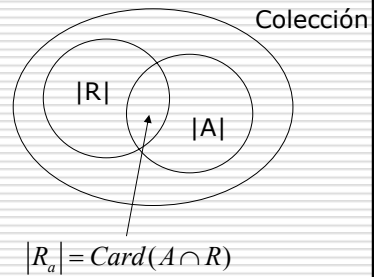
- ❑ Objetivo: evaluar la funcionalidad y eficiencia de un sistema de IR
- ❑ Medidas usuales de evaluación:
  - Tiempo de respuesta
  - Espacio utilizado
  - Evaluación de la eficiencia de recuperación basada en
    - ❑ *Colecciones prueba* (TIPSTER, CACM, CISI,...)
    - ❑ *Medidas de evaluación*

# Medidas de evaluación

- El usuario desea realizar una solicitud de información sobre una colección de docs
- Sean
  - R: docs relevantes
  - A: docs respuesta

$$Recall = \frac{|R_a|}{|R|}$$

$$Precision = \frac{|R_a|}{|A|}$$



Almacenamiento y recuperación de  
información en la Web

## Evolución recuperación - precisión

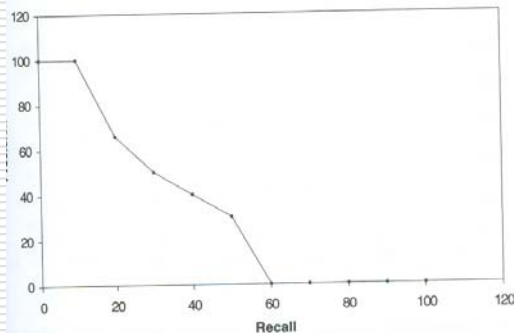


Figure 3.2 Precision at 11 standard recall levels.

- Usualmente se usan 11 niveles de recuperación

Almacenamiento y recuperación de  
información en la Web

## Evolución recuperación - precisión

1	d <sub>123</sub>	6	d <sub>9</sub>	11	d <sub>38</sub>
2	d <sub>84</sub>	7	d <sub>511</sub>	12	d <sub>48</sub>
3	d <sub>56</sub>	8	d <sub>129</sub>	13	d <sub>250</sub>
4	d <sub>6</sub>	9	d <sub>187</sub>	14	d <sub>113</sub>
5	d <sub>8</sub>	10	d <sub>25</sub>	15	d <sub>3</sub>

$R_q = \{d_3, d_{56}, d_{129}\}$  documentos relevantes

---

$A = \{d_{123}\} \Rightarrow |A| = 1$

$\text{Recall} = \frac{0}{3} = 0$      $\text{Precision} = \frac{0}{1} = 0$

---

$A = \{d_{123}, d_{84}\} \Rightarrow |A| = 2$

$\text{Recall} = \frac{0}{3} = 0$      $\text{Precision} = \frac{0}{2} = 0$

---

$A = \{d_{123}, d_{84}, d_{56}\} \Rightarrow |A| = 3$

$\text{Recall} = \frac{1}{3} = 0,33\%$      $\text{Precision} = \frac{1}{3} = 0,33\%$

Almacenamiento y recuperación de  
información en la Web

## Evolución recuperación - precisión

Doc	A	R <sub>a</sub>	R	Recall	Precision
1	1	0	3	0	0
2	2	0	3	0	0
3	3	1	3	33,3	33,3
4	4	1	3	33,3	25
5	5	1	3	33,3	20
6	6	1	3	33,3	16,6
7	7	1	3	33,3	14,2
8	8	2	3	66,6	25
9	9	2	3	66,6	22,2

Almacenamiento y recuperación de  
información en la Web

## Evolución recuperación - precisión

Doc	A	R <sub>a</sub>	R	Recall	Precision
10	10	2	3	66,6	20
11	11	2	3	66,6	18,1
12	12	2	3	66,6	16,6
13	13	2	3	66,6	15,3
14	14	2	3	66,6	14,2
15	15	3	3	100	20

Almacenamiento y recuperación de  
información en la Web

## Precisión promedio

- ❑ Los algoritmos de recuperación evalúan qs diferentes

$$\bar{P}(r) = \frac{1}{N_q} \sum_{i=1}^{N_q} P_i(r)$$

- ❑ Para evaluar un algoritmo se promedian las precisiones en cada nivel de recuperación para todas las qs

$\bar{P}(r)$     *Precision media al nivel r de recuperacion*

$P_i(r)$     *Precision al nivel r para la i-esima q*

$N_q$     *Nº de qs utilizadas*

Almacenamiento y recuperación de  
información en la Web

## Interpolación de la Precisión Promedio

- Para representar la evolución de la precisión promedio se efectúa una interpolación de las precisiones en cada nivel de recuperación

$$P(r) = \max_{r_j \leq r \leq r_{j+1}} P(r)$$

$$r_j \in \{0, 1, 2, \dots, 10\}$$

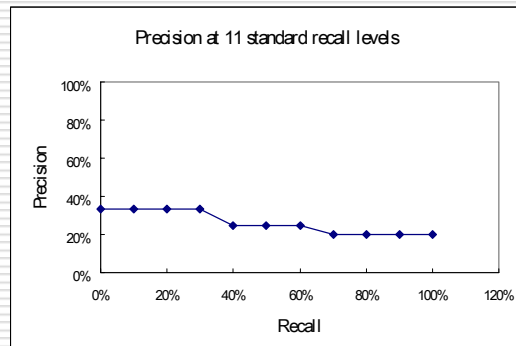
Almacenamiento y recuperación de información en la Web

## Interpolación de la Precisión Promedio

- $R_q = \{d_3, d_{56}, d_{129}\}$
- $A_q = \{d_{123}, d_{84}, \mathbf{d_{56}}, d_6, d_8, d_9, d_{511}, \mathbf{d_{129}}, d_{187}, d_{25}, d_{38}, d_{48}, d_{250}, d_{113}, \mathbf{d_3}\}$
- R=33%, P=33%
- R=66%, P=25%
- R=100%, P=20%

$$P(r) = \max_{r_j \leq r \leq r_{j+1}} P(r)$$

$$r_j \in \{0, 1, 2, \dots, 10\}$$



Almacenamiento y recuperación de información en la Web

## Precisión Promedio en $n$

---

- ❑ Otra medida usual es el cálculo de la precisión promedio tras  $n$  documentos relevantes recuperados (p. ej. 5, 10, 20, 30, 50, 100)
- ❑ Se calcula la media de las precisiones obtenidas hasta el nivel de corte
- ❑ Este sistema favorece a los buscadores que recuperan los documentos relevantes rápido
- ❑ Ej: si al nivel 5 tenemos unas medidas de precisión de 1, 0.66, 0.5, 0.4, 0.3
  - $AvP = (1 + .66 + .5 + .4 + .3) / 5 = 0.572$

---

Almacenamiento y recuperación de  
información en la Web

## R-Precision

---

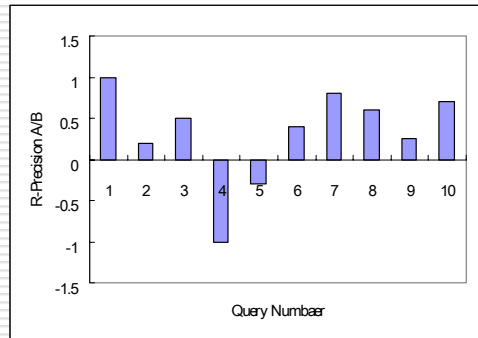
- ❑ La idea es generar un valor resumen del ranking mediante la precisión en la posición R-ésima del ranking, siendo R el nº total de docs relevantes
  - Ej: la 3-Precision es: 0.33
- ❑ Es útil para comprobar el comportamiento de un algoritmo frente a cada ítem

---

Almacenamiento y recuperación de  
información en la Web

# Histogramas de R-Precisión

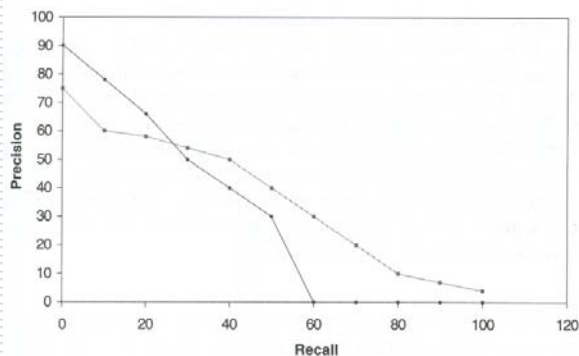
- Las medidas de R-Precisión se pueden usar para comparar el comportamiento de dos algoritmos de forma gráfica a lo largo de diferentes búsquedas



■  $RP_{A/B}(i) = RP_A(i) - RP_B(i)$

Almacenamiento y recuperación de  
información en la Web

# Comparación de algoritmos IR



3.4 Average recall versus precision figures for two distinct retrieval algo-

Almacenamiento y recuperación de  
información en la Web

## Consideraciones generales sobre las medidas r y P

- ❑ En colecciones grandes no es posible disponer de un conocimiento detallado de los docs
- ❑ Considerar la combinación de ambas medidas (r y P)
- ❑ Las medidas de r y P son para qs procesadas en *batch*. Para sistemas de IR interactivos son más convenientes medidas que cuantifiquen la bondad del proceso

Almacenamiento y recuperación de  
información en la Web

## Medidas alternativas, I

- ❑ Media armónica
  - Combina r y P
  - $F=0$  no se recuperan docs relevantes
  - $F=1$  todos los docs recuperados son relevantes
  - r y P altas  $\rightarrow$  F alta

$$F(j) = \frac{2}{\frac{1}{r(j)} + \frac{1}{P(j)}} \quad F \in [0,1]$$

$r(j)$  recuperación j-esimo doc  
 $P(j)$  precision j-esimo doc  
 $F(j)$  recuperación j-esimo doc

Almacenamiento y recuperación de  
información en la Web



## Medidas alternativas, II

### □ Medida E (de *evaluación*)

- Combina r y P
- $b=1 \rightarrow E(j)=1-F(j)$
- $b>1$ 
  - usuario interesado en P
- $b<1$ 
  - usuario interesado en r

$$E(j) = 1 - \frac{1+b^2}{\frac{b^2}{r(j)} + \frac{1}{P(j)}}$$

$r(j)$  recuperación j-esimo doc

$P(j)$  precision j-esimo doc

$F(j)$  recuperación j-esimo doc

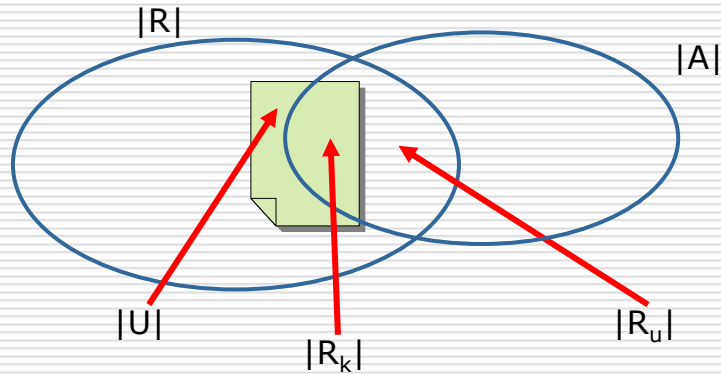
$b$  parametro definido por el usuario

## Medidas alternativas, III (orientadas al usuario)

- Pretenden tener en cuenta las diferencias existentes entre usuarios interesados por un doc
- Contexto
  - C: Colección de docs de referencia
  - I: Ejemplo de solicitud de info
  - R: Conjunto relevante de docs para I
  - A: Conjunto recuperado
  - U: Subconjunto de R conocido por el usuario
    - $|U| = \text{Card}(U)$
  - $A \cap U$ : docs conocidos por el usuario relevantes y recuperados
    - $|R_k| = \text{Card}(A \cap U)$
    - $|R_u|$ 
      - nº de docs relevantes desconocidos por el usuario que fueron recuperados

## Medidas alternativas, IV (orientadas al usuario)

---



---

Almacenamiento y recuperación de  
información en la Web

## Medidas alternativas, V (orientadas al usuario)

---

### ☐ Alcance

- Fracción de los docs conocidos relevantes recuperados

$$alcance = \frac{|R_k|}{|U|}$$

### ☐ Novedad

- Fracción de los docs desconocidos relevantes recuperados

$$novedad = \frac{|R_u|}{|R_u| + |R_k|}$$

---

Almacenamiento y recuperación de  
información en la Web

## Medidas alternativas, VI (orientadas al usuario)

---

### ☐ Recuperación relativa

- Cociente entre el nº de docs relevantes encontrados y el nº de docs relevantes que el usuario esperaba encontrar
  - ☐ Si encuentra tantos como esperaba -->  $RR=1$

### ☐ Esfuerzo de recuperación

- Cociente entre el nº de docs relevantes que el usuario esperaba encontrar y el nº de docs examinados con el fin de cubrir el nº anterior

---

Almacenamiento y recuperación de  
información en la Web

## Colecciones, I

---

### ☐ TIPSTER/TREC

- TREC  $\equiv$  Text Retrieval Conference (1990)
  - ☐ Dedicada a experimentación con colecciones grandes ( $\approx 1.000.000$  docs)
  - ☐ Colección TREC: 6 CDs  $\approx 1\text{Gb}$  cada uno
  - ☐ Docs de: WSJ, AP, FT, etc.
  - ☐ <http://trec.nist.gov/>

---

Almacenamiento y recuperación de  
información en la Web

## Colecciones, II

---

### □ CACM

- 3204 artículos de *Communications of the ACM* (1958-1979)
  - Campos
    - Autores, fecha edición, palabras clave (reducidas a su raíz gramatical) de título y abstract, referencias entre artículos, bibliografía, etc.
  - Incluye un conjunto de 52 solicitudes de información. Ej: "Qué artículos hay que traten de TSS (Time Sharing System), sistema operativo de ordenadores IBM"
  - El nº medio de docs relevantes para cada I es pequeño, en torno a 15.

---

Almacenamiento y recuperación de  
información en la Web

## Colecciones, III

---

### □ ISI (o CISI)

- 1460 docs escogidos del ISI (Institute of Scientific Information)
  - Los docs escogidos se seleccionaron como los más citados en un estudio sobre citación realizado por Small
  - Propósito general: facilitar la investigación sobre similitudes basadas en términos y patrones de referencias cruzadas
  - Campos
    - Autores, palabras clave (reducidas a su raíz gramatical) de título y abstract y nº de "cocitaciones" para cada par de artículos
  - Incluye un conjunto de 35 solicitudes de información en LN y qs booleanas y 41 sólo en LN.
    - El nº medio de docs relevantes para cada I es grande, en torno a 50.

---

Almacenamiento y recuperación de  
información en la Web

# Calidad de los resultados

---

- ❑ ¿Se pueden aplicar los criterios de medida de la IR clásica a la *web*?
- ❑ En IR clásica las medidas usadas son:
  - Precisión: % de páginas recuperadas que son relevantes
  - Recuperación: % de páginas relevantes que son recuperadas
- ❑ En *web* IR:
  - El término *relevante* se liga al de *calidad*
  - Una página es valorable si es una página de calidad para el objeto de la búsqueda
  - Precisión: número de páginas valables recuperadas