

# **Sistemi per il recupero delle informazioni**

**Gabriele Pozzani**

**A.A. 2012/2013**

**Corso di Laurea Magistrale in  
Editoria e Giornalismo**

**SRI: valutazione delle prestazioni**

## **Efficienza vs Efficacia**

- In un SRI si possono misurare
  - Efficienza: come il sistema si comporta in termini di tempi di risposta, uso della memoria, ecc...
  - Efficacia: quanto il sistema è in grado di soddisfare l'utente fornendogli le (sole) informazioni rilevanti
    - Include sia aspetti quantitativi che soggettivi

3

## **Efficacia, valutazione delle performance**

- Al fine di misurare l'efficacia di un SRI sono state sviluppate diverse misure numeriche basate sui risultati di un SRI
  - Associate alla rilevanza dei risultati per l'utente
  - Basate sul confronto dei risultati di un SRI con risultati proposti dagli utenti

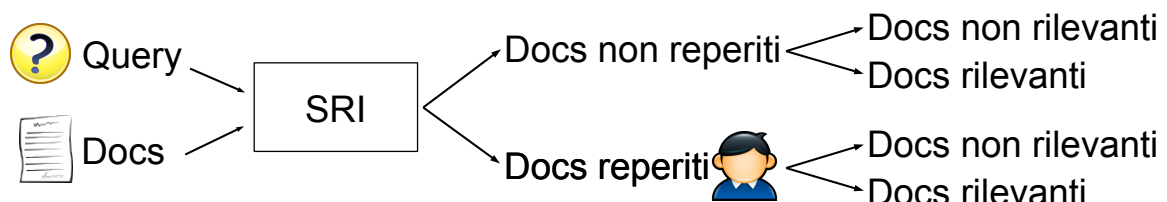
4

## Misura delle performance

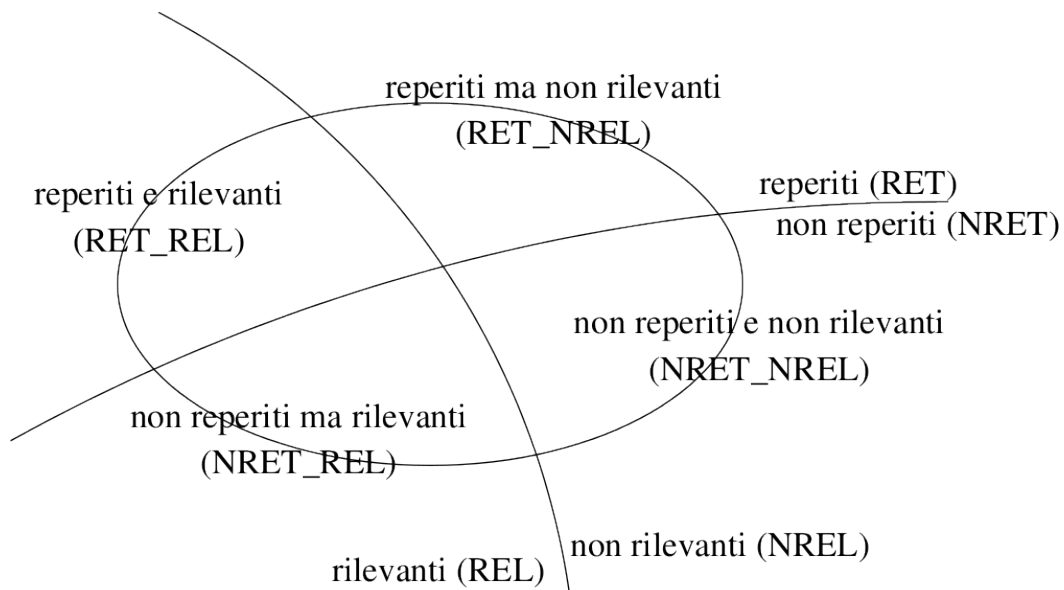
- Vi sono due grandi classi di misure per la valutazione dei SRI
  - Basate sul partizionamento dello spazio dei documenti
  - Orientate all'utente

## Partizionamento dello spazio dei documenti (I)

- Le misure di efficacia di un SRI sono basate sul partizionamento dello spazio dei documenti
- In seguito ad una query, i documenti si possono dividere tramite due classificazioni indipendenti ed ortogonali
  - rilevanti (REL) vs non rilevanti (NREL)
  - reperiti (RET) vs non reperiti (NRET)



## Partizionamento dello spazio dei documenti (II)



7

## Partizionamento dello spazio dei documenti (III)

- RET\_REL: documenti che sono rilevanti per l'utente e vengono correttamente identificati come tali e ritornati;
- RET\_NREL: documenti che sono ritornati dal sistema di IR nonostante non siano rilevanti
  - Costituiscono "rumore"
  - Da minimizzare
  - Detti anche *false hits*
- NRET\_REL: documenti che pur essendo rilevanti non vengono riconosciuti come tali e quindi non vengono ritornati
  - i documenti per cui il sistema è "silenzioso"
  - Da minimizzare
  - Detti anche *false dismissals*
- NRET\_NREL: documenti che non sono rilevanti e che il sistema identifica correttamente come tali e quindi non ritorna
  - documenti giustamente da omettere (e omessi) nel risultato

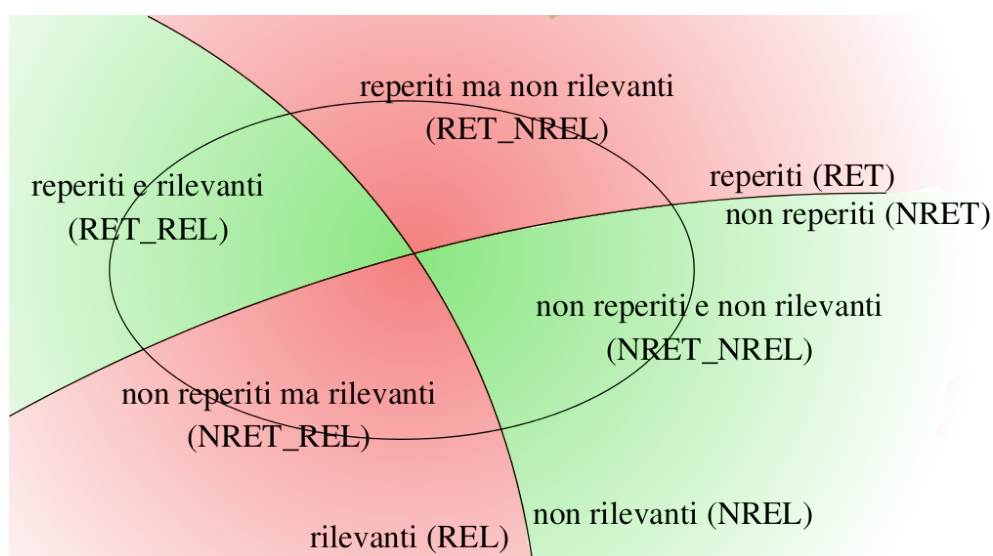
8

## Partizionamento dello spazio dei documenti (IV)

	Doc. Rilevanti (REL)	Doc. non rilevanti (NREL)
Doc. reperiti (RET)	Corretti	Inesatti
Doc. non reperiti (NRET)	Omessi	Da omettere

9

## Partizionamento dello spazio dei documenti (V)



Sistema ideale: **nessun false hits**  
**nessun false dismissals**

10

## Precisione e Richiamo

- Negli anni '50 Cyril Cleverdon condusse i primi esperimenti sui sistemi di indicizzazione e IR
- Quegli esperimenti portarono alla definizione alle moderne misure di efficacia di un SRI:
  - Richiamo (recall)
  - Precisione (precision)

11

## Precisione e Richiamo

- Richiamo (recall): la frazione di documenti rilevanti che sono recuperati
  - $R = \frac{\#RET\_REL}{\#REL}$
  - Valuta la capacità di trovare i doc. rilevanti
- Precisione (precision): la frazione di documenti ritornati che sono rilevanti
  - $P = \frac{\#RET\_REL}{\#RET}$
  - Valuta la capacità di rigettare i doc. non rilevanti

12

## Precisione e Richiamo

- La precisione è calcolabile in modo preciso a partire dal risultato di una query
  - si conoscono i valori esatti sia di #RET\_REL che di #RET
- Il richiamo è una misura “inesatta” non calcolabile precisamente
  - Richiede di conoscere #REL, il numero di docs rilevanti in tutta la collezione
    - Molto difficile o impossibile

13

## Richiamo

- Entrambe le misure assumono valori tra 0 e 1
  - 0: caso pessimo
  - 1: caso ottimo
- Richiamo:
  - Se vale 1 il sistema è stato in grado di recuperare tutti i documenti rilevanti
    - non vi sono false dismissals
  - Se vale 0 (o un valore molto vicino allo 0) il sistema ha recuperato nessuno (pochi) dei documenti rilevanti

14

# Precisione

- Precisione:
  - Se vale 1 il sistema è stato in grado di recuperare tutti e solo documenti rilevanti
    - non vi sono false hits
  - Se vale 0 (o un valore molto vicino allo 0) nessuno (pochi) docs recuperati sono rilevanti

15

## Uso di precisione e richiamo

- Sono utilizzate per misurare e confrontare l'efficacia di diversi SRI
- Per ogni SRI vengono eseguite un insieme prestabilito di interrogazioni e con il risultato di ciascuna si calcolano precisione e richiamo.
  - Confrontando i valori delle due misure su queste interrogazioni "campione" dei due SRI è possibile, ma non sempre, osservare se uno dei due sistemi è "migliore" dell'altro

16



## Confronto tra SRI: esempio (1)

- Si consideri il SRI  $IRS_1$  e una query  $Q_1$
- Sia  $R_{Q_1} = \{d_3, d_5, d_9, d_{25}, d_{39}, d_{44}, d_{56}, d_{71}, d_{89}, d_{123}\}$  l'insieme dei doc rilevanti per  $Q_1$
- Poniamo che  $IRS_1$  in seguito a  $Q_1$  ritorni i seguenti documenti nel seguente ordine
  - I punti indicano i documenti rilevanti

01. $d_{123}$ •	06. $d_9$ •	11. $d_{38}$
02. $d_{84}$	07. $d_{511}$	12. $d_{48}$
03. $d_{56}$ •	08. $d_{129}$	13. $d_{250}$
04. $d_6$	09. $d_{187}$	14. $d_{113}$
05. $d_8$	10. $d_{25}$ •	15. $d_3$ •

17

## Confronto tra SRI: esempio (2)

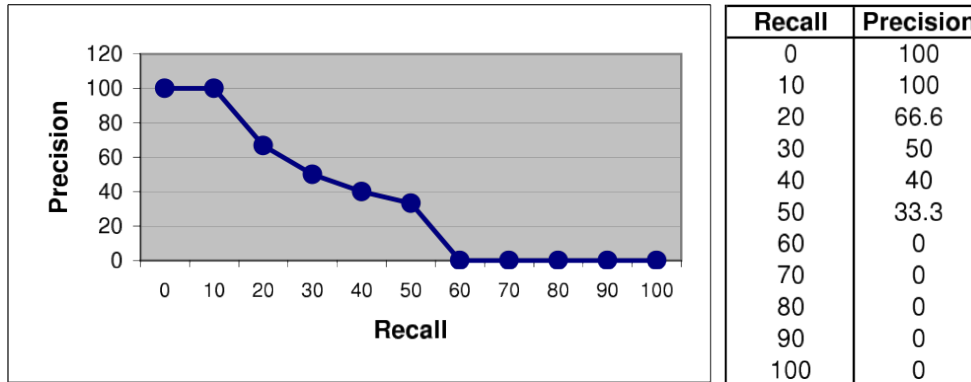
- Il doc  $d_{123}$  è al primo posto ed è rilevante
  - Esso è il 10% dei doc rilevanti
  - Si ha una precisione del 100% e un richiamo del 10%
- Il successivo doc rilevante è  $d_{56}$ , al terzo posto
  - 2 doc su 3 sono rilevanti e 2 doc rilevanti su 10 sono stati tornati
  - Si ha una precisione del 66,6% e un richiamo del 20%

01. $d_{123}$ •	06. $d_9$ •	11. $d_{38}$
02. $d_{84}$	07. $d_{511}$	12. $d_{48}$
03. $d_{56}$ •	08. $d_{129}$	13. $d_{250}$
04. $d_6$	09. $d_{187}$	14. $d_{113}$
05. $d_8$	10. $d_{25}$ •	15. $d_3$ •

18

## Confronto tra SRI: esempio (3)

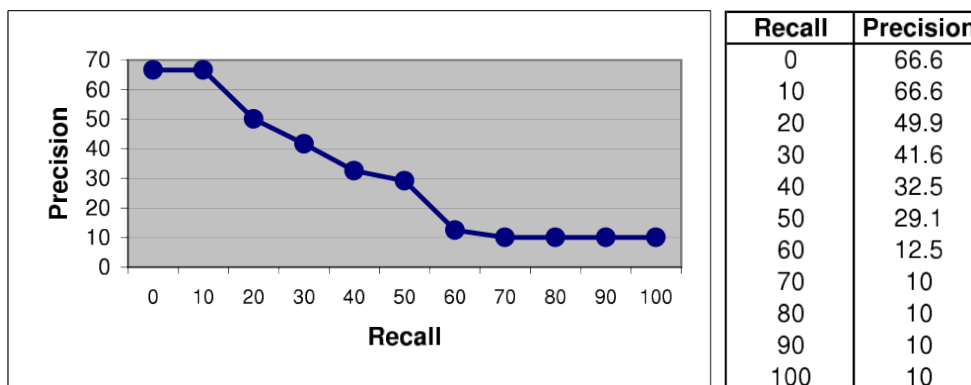
- Proseguendo così possiamo disegnare un grafico precisione-richiamo



19

## Confronto tra SRI: esempio (4)

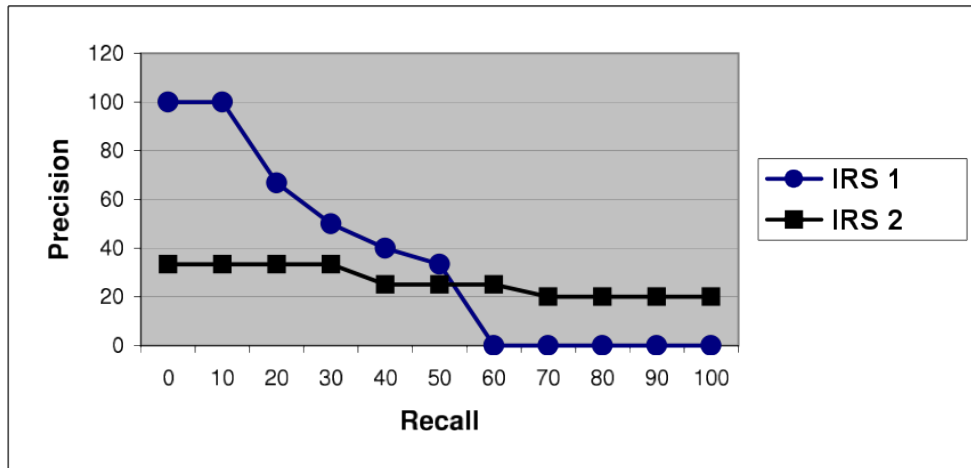
- Sullo stesso SRI si può rieseguire questo procedimento con diverse altre query
  - Per ogni livello di richiamo di fa la media delle precisioni ottenute con le diverse query
  - Si ottiene un grafico precisione-richiamo generale



20

## Confronto tra SRI: esempio (5)

- Facendo la stessa cosa per un secondo SRI  $IRS_2$  e sovrapponendo i due grafici si possono confrontare i due SRI



21

## Relazione tra precisione e richiamo (1)

- vi è una legge
  - non è dimostrabile essere sempre vero
  - è vero in media, nella maggior parte dei casi
- inversa tra precisione e richiamo
  - All'aumentare della precisione cala il richiamo e viceversa
- Lo si vede esaminando i seguenti due casi "base"

22

## Relazione tra precisione e richiamo (2)

- Poniamo che il sistema ritorni esattamente un solo documento tra tutti quelli della collezione e che questo documento sia rilevante
  - la precisione vale 1
  - il richiamo, anche se non possiamo calcolarlo esattamente, possiamo dire che sia basso, potendo supporre che i documenti rilevanti siano (molti) più di 1 soltanto.
- Poniamo ora che il sistema ritorni tutti i documenti della collezione
  - il richiamo varrà 1
  - la precisione sarà bassa, in quanto tra tutti i documenti tornati vi saranno anche tutti quelli non rilevanti.

23

## Precisione vs Richiamo

- Il sistema ideale non esiste
- Meglio preferire un'alta precisione o un alto richiamo??
  - Cranfield sostenne che nella maggior parte dei casi non è necessario un alto richiamo
  - Cleverdon nel 1991 ha suggerito che gli utenti preferiscano un'alta precisione ad un alto richiamo
  - Su nel 1994 ha invece suggerito che nessuna delle due misure sia significativa per l'utente

24

## Precisione e richiamo: conclusioni

- Precisione e richiamo sono state ampiamente usate
- Ma hanno alcuni problemi
  - 1) La stima del richiamo non è semplice, richiedendo un'approfondita conoscenza di tutti i doc nella collezione
  - 2) In alcuni casi l'uso di più misure può essere problematico
  - 3) Non tengono conto della possibilità da parte dell'utente di raffinare la ricerca in passi successivi
- I grafici precisione-richiamo sono il metodo standard per valutare i SRI

25

## Fallout e Generalità

- Altre misure di efficacia di un SRI sono
  - Fallout: porzione di doc non rilevanti che vengono ritornati

$$F = \frac{\#RET\_NREL}{\#NREL}$$

- Generalità: porzione di doc rilevanti sul totale dei documenti nella collezione

$$G = \frac{\#REL}{N}$$

26

## Legame tra le quattro misure (1)

- Precisione, richiamo, fallout e generalità sono legate dalla seguente equazione

$$\frac{R}{F} = \frac{\frac{P}{1-P}}{\frac{G}{1-G}}$$

- $\frac{P}{1-P}$  : rapporto tra il numero di doc rilevanti recuperati e il numero di doc non rilevanti recuperati
- $\frac{G}{1-G}$  : rapporto tra il numero di doc rilevanti e di doc non rilevanti nella collezione
- $\frac{R}{F}$  : rapporto tra le prestazioni del sistema su doc rilevanti e le prestazioni su documenti non rilevanti

27

## Legame tra le quattro misure (2)

- $\frac{P}{1-P}$  : meglio valori alti
- $\frac{G}{1-G}$  : meglio valori alti
- $\frac{R}{F}$  :  $>1$  se il sistema è migliore nel recuperare i doc rilevanti di quanto non lo sia nell'evitare i doc non rilevanti

Misura che cerca di racchiudere tutte le altre

28

## Misure orientate all'utente (1)

- Le misure precedenti partono dal presupposto che gli utenti siano un gruppo omogeneo
  - Data una query e una risposta ad essa tutti gli utenti valutino la qualità del risultato allo stesso modo
  - Non è proprio così
    - La qualità di un risultato di un SRI è soggettiva
- Sono state definite le misure orientate all'utente

Utenti diversi valutano diversamente la rilevanza dei risultati

29

## Misure orientate all'utente (2)

- Rapporto di copertura
  - Rapporto tra i documenti rilevanti noti all'utente che sono recuperati e il totale dei doc nella collezione noti all'utente
- Rapporto di novità
  - Rapporto tra i documenti rilevanti che sono recuperati e che non erano noti all'utente e i documenti rilevanti recuperati
- Richiamo relativo
  - Rapporto tra il numero di doc rilevanti recuperati esaminati dall'utente e il numero di doc rilevanti che l'utente avrebbe voluto esaminare
- Lunghezza di ricerca attesa
  - Numero medio di documenti che l'utente deve esaminare per trovare il numero di doc rilevanti che desidera

30

## Misure orientate all'utente: esempio

- Poniamo che
  - All'utente siano già noti 5 documenti
  - L'utente desideri 10 documenti rilevanti
  - Il SRI recuperi 10 documenti
    - 2 non rilevanti
    - 8 rilevanti
      - 3 noti all'utente
      - 5 non noti all'utente
- Rapporto di copertura =  $3/5$
- Rapporto di novità =  $5/8$
- Richiamo relativo =  $8/10$

31

## Lunghezza di ricerca attesa (1)

- Numero medio di documenti che l'utente deve esaminare per trovare il numero di doc rilevanti che desidera **Per l'utente**
- I documenti recuperati vengano divisi in  $k$  pagine  $S_1, S_2, \dots, S_k$ 
  - I documenti in ogni pagina sono equivalenti rispetto al criterio di recupero
  - I documenti in ogni pagina sono migliori di quelli nella pagina successiva **Per il SRI**
- Ad esempio, data l'interrogazione  $t_1$  OR  $t_2$  OR ... OR  $t_n$ 
  - $S_1$  contiene i documenti che contengono  $n$  termini
  - $S_2$  contiene i documenti che contengono  $n-1$  termini
  - ...

**Per il SRI hanno la stessa rilevanza**

**Per il SRI**

**Nota (sempre vera): vi è una discrepanza tra ciò che il SRI pensa rilevante e ritorna e ciò che è (idealmente) rilevante per l'utente**

32



## Lunghezza di ricerca attesa (2)

- L'utente per recuperare i documenti che ritiene rilevanti può dover esaminare diverse pagine
- Quanti documenti in queste pagine deve esaminare per trovare il numero di documenti rilevanti che desidera?
  - Lunghezza di ricerca attesa

33

## Lunghezza di ricerca attesa: esempio (1)

- I doc recuperati sono suddivisi in 3 pagine
  - $S_1$  contiene 3 doc
    - 1 rilevante
    - 2 non rilevanti
  - $S_2$  contiene 5 doc
    - 4 rilevanti
    - 1 non rilevante
  - $S_3$  contiene 5 documenti
    - 2 rilevanti
    - 3 non rilevanti

34

## Lunghezza di ricerca attesa: esempio (2)

- Se l'utente vuole 1 solo documento rilevante
  - Deve esaminare solo  $S_1$  che però potrebbero presentarsi in diversi ordini
    - R NR NR
    - NR R NR
    - NR NR R
  - Il doc rilevante potrebbe essere il primo esaminato, il secondo o il terzo

35

## Lunghezza di ricerca attesa: esempio (3)

- I 3 ordini hanno la stessa probabilità:  $1/3$
- L'utente ha prob.  $1/3$  di esaminare un solo doc
  - $1/3$  di esaminare 2 doc
  - $1/3$  di esaminare 3 doc
- La lunghezza di ricerca attesa per trovare un doc rilevante è:  
$$1 \times \frac{1}{3} + 2 \times \frac{1}{3} + 3 \times \frac{1}{3} = 2$$
  - L'utente in media deve esaminare 2 doc per trovarne 1 rilevante

36

## Lunghezza di ricerca attesa: altro esempio (1)

- Se l'utente vuole 6 documenti rilevanti
  - In  $S_1$  e  $S_2$  ne trova 5 esaminando 8 documenti
  - Il 6° doc rilevante sarà in  $S_3$
  - I possibili ordinamenti di  $S_3$  sono
    1. R R NR NR NR
    2. R NR R NR NR
    3. R NR NR R NR
    4. R NR NR NR R
    5. NR R R NR NR
    6. NR R NR R NR
    7. NR R NR NR R
    8. NR NR R R NR
    9. NR NR R NR R
    10. NR NR NR R R

37

## Lunghezza di ricerca attesa: altro esempio (2)

- 10 ordinamenti possibili
  - Nei primi 4 l'utente esamina 1 doc
  - Nei successivi 3 l'utente esamina 2 doc
  - Nei successivi 2 l'utente esamina 3 doc
  - Nell'ultimo l'utente esamina 4 doc

– Lunghezza di ricerca attesa

$$8 + \left(1 \times \frac{4}{10} + 2 \times \frac{3}{10} + 3 \times \frac{2}{10} + 4 \times \frac{1}{10}\right) = 10$$

38