

Biblioteca Federata di Medicina "Ferdinando Rossi"



UNIVERSITÀ
DEGLI STUDI
DI TORINO

Dallo scenario al quesito di ricerca: tecniche di IR

a cura di
Nicoletta Colombi
Responsabile U.O. 3 Comunicazione Expert
searching e bibliometria

Quest'opera è stata rilasciata con licenza Creative Commons. Attribuzione - Non commerciale - Condividi allo stesso modo 4.0 Internazionale

Aggiornato al 13 ottobre 2016
Versione 1.0



Information retrieval (IR)

L'insieme delle tecniche utilizzate per il recupero mirato dell'informazione in formato elettronico.

L'utente digita una query ("interrogazione"), ovvero una stringa di parole-chiave rappresentanti l'informazione richiesta.

Essa viene messa in relazione con gli oggetti presenti nell'archivio di dati consultato.

Il sistema fornisce come risposta un insieme di record che soddisfano le condizioni richieste.

Dallo scenario al quesito di ricerca

Scenario = l'esigenza conoscitiva alla base della ricerca.

PER OTTENERE DELLE RISPOSTE DALLA BD
OCCORRE TRASFORMARLO IN

Quesito (Query) = domanda precisa da porre alla BD
in un linguaggio ad essa comprensibile
(**parole chiave + operatori logici**)

NB: Uno scenario può essere composto da più quesiti.



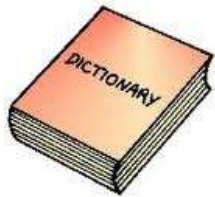
IL QUESITO DI RICERCA (QUERY)

Il quesito di ricerca



Che cosa sto cercando?

I concetti fondamentali individuabili all'interno dello scenario.



Con quali parole?

I termini che li esprimono. Sinonimi e contrari. Parole libere e, nel caso delle BD bibliografiche, termini di Tesaurus.



In quale relazione tra loro?

Le relazioni logico-astratte che legano tali termini (operatori booleani).

Che cosa sto cercando?

Quesiti “Background”

- In campo clinico-assistenziale, sono i quesiti **di carattere generale** riguardo una patologia, un trattamento, ecc. Mirano a recuperare **conoscenze di base**: cos'è una determinata malattia? Quali sono le sue cause? Come si presenta? Che tipo di terapie le si possono applicare?

Quesiti “Foreground”

- In campo clinico-assistenziale, sono quesiti **specifici**, fondati su conoscenze di base già acquisite. Spesso sono finalizzati a indagare un **confronto di efficacia** tra possibili alternative diagnostiche o terapeutiche percorribili. Su essi si basa la pratica dell'**EBM**.

Un metodo di strutturazione del quesito background

1 Trasformare il tema oggetto del proprio interesse conoscitivo in una domanda.

Es.: Voglio esplorare le ricadute psicologiche e le difficoltà di vita domestica e quotidiana vissute dai famigliari di pazienti schizofrenici. -> **Quali problematiche si trovano ad affrontare i famigliari di pazienti schizofrenici?**

2 Scrivere il titolo della pubblicazione in grado idealmente di rispondere a questa domanda.

Es.: **Le problematiche dei famigliari di pazienti schizofrenici**

3 Separare i principali concetti che lo compongono con delle linee.

Es.: Le problematiche dei familiari di pazienti schizofrenici

4 Elencare, sotto ciascuno di esso, possibili sinonimi, varianti concettuali, antonimi, termini tecnici, ecc.

Es.: Le problematiche dei familiari di pazienti schizofrenici

ardello	genitori	schizofrenia
qualità della vita	parenti	psicosi
benessere	coniugi	...
...	...	

5 Tradurre l'operazione anche in inglese.

La strutturazione del quesito foreground: il P.I.C.O.

Paziente/popolazione/problema (Patient/Population/Problem):

Come meglio descriverei il mio paziente, la popolazione o il problema di cui mi sto occupando?

Intervento/esposizione (Intervention/exposure):

Che tipo di intervento (es.: un trattamento), fattore prognostico, test diagnostico o esposizione sto considerando?

Comparazione/controllo (Comparison/Control):

Con quale alternativa posso confrontarlo? [OPZIONALE]

Outcome (esito):

Che cosa spero di ottenere, migliorare, misurare o influenzare?

Diagnosis Questions: P.I.C.O.

P La patologia o condizione clinica che si vuole identificare (con attenzione alle caratteristiche specifiche del paziente)

I L'esame/procedura diagnostica che si desidera applicare

[C] L'esame/procedura diagnostica "gold standard" (se esiste) o un altro esame diagnostico

O Misure che dimostrino l'utilità diagnostica dell'esame in "I".
Es.: sensibilità, specificità, valore positivo/negativo di previsione...

Therapy Questions: P.I.C.O.

P La patologia o la condizione del paziente

I La specifica misura terapeutica di interesse. Es.: terapia, medicazione, intervento chirurgico, modifica dello stile di vita...

[C] Un altro trattamento o intervento, oppure un placebo

O Risultati clinici misurabili importanti per il paziente. Es.: tasso di mortalità, dolore, giorni di lavoro persi, recidive...

P: "Come descriverei un gruppo di pazienti simili al mio?
Bilanciare la precisione con la brevità.

I: Quale intervento principale sto prendendo in considerazione?
Essere specifici.

C (se necessario): Qual è la principale alternativa da comparare all'intervento prescelto? **Di nuovo, essere specifici.**

O: Che cosa spero di ottenere? Che cosa questo intervento può realmente produrre? **Di nuovo, essere specifici."**

D.L. Sackett et al., *La medicina basata sull'evidenza. Come praticare ed insegnare la EBM*, Torino: Centro Scientifico Editore 1998, p. 24.

Un esempio di quesito foreground

L'inserimento di un familiare di paziente schizofrenico all'interno di un gruppo di auto-mutuo aiuto costituisce una forma efficace di intervento di supporto psico-sociale?

P Familiare/parente/caregiver di paziente schizofrenico/schizofrenia

I Gruppo di auto-mutuo aiuto

[C] [Psicoterapia convenzionale/nessun intervento...]

O Miglioramento qualità della vita/riduzione stress/depressione...

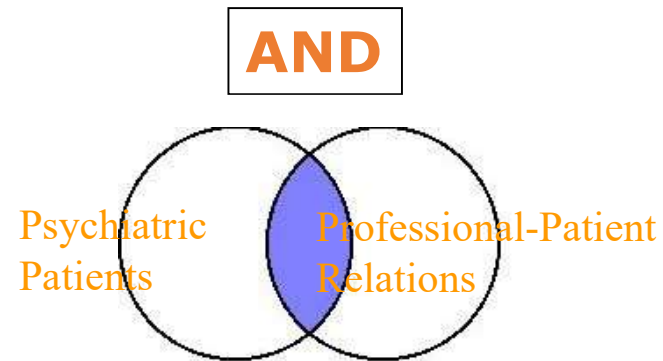
Relazioni logico-astratte tra concetti: gli operatori booleani

Operatori logici definiti verso la metà del XIX secolo
da George Boole, matematico dell'University College di Cork.

Nell'IR permettono di incrociare i diversi concetti
realizzando le operazioni insiemistiche
dell'intersezione, dell'unione e della complementazione (o differenza).

Wikipedia, *Algebra di Boole*, [http://it.wikipedia.org/wiki/Algebra di Boole](http://it.wikipedia.org/wiki/Algebra_di_Boole)

Operatori booleani: AND



Intersezione: RESTRINGE

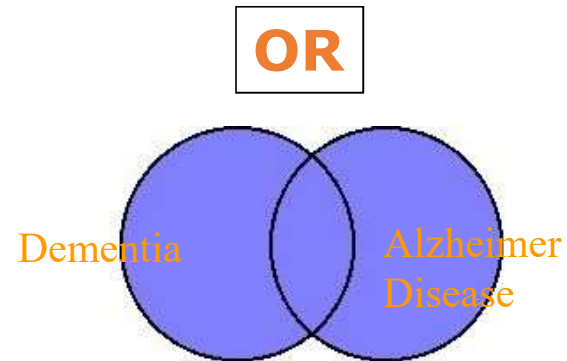
Psychiatric Patients AND Professional-Patient Relations

dà come risultato tutti i lavori che trattano i due argomenti in maniera congiunta.

Esclude i documenti in cui essi non sono compresenti.

Si utilizza pertanto per delimitare il campo di ricerca dettagliando il quesito nei suoi parametri essenziali.

Operatori booleani: OR



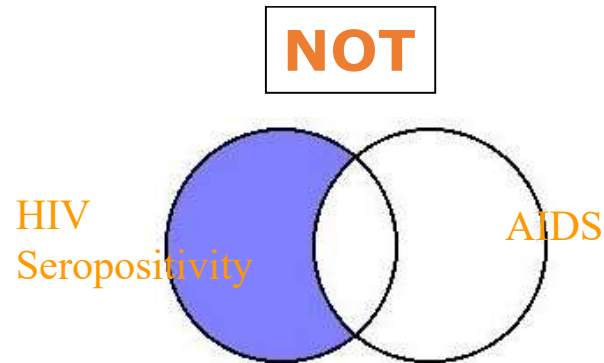
Unione: ALLARGA

Dementia OR Alzheimer Disease

dà come risultato tutti i lavori che trattano i due argomenti sia singolarmente che in compresenza tra loro.

Si utilizza pertanto per allargare il campo di ricerca contemplando possibili sinonimi o concetti affini.

Operatori booleani: NOT



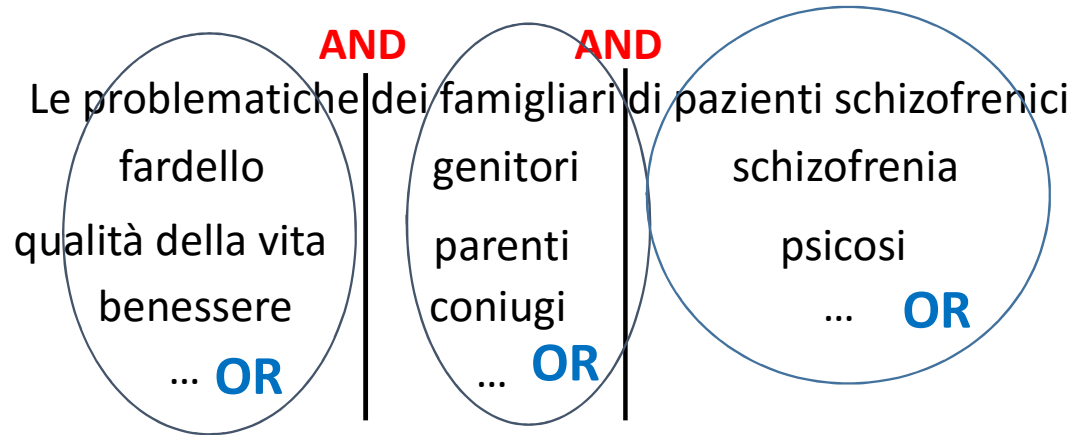
Differenza: ELIMINA

HIV Seropositivity NOT AIDS

dà come risultato tutti i lavori che trattano di sieropositività in assenza di AIDS conclamato (concetto pertinente, ma in questo caso non di interesse).

Si utilizza pertanto per limare ulteriormente lo scenario di ricerca (spesso a posteriori).

Quindi...



P (Famigliare OR parente OR caregiver) AND (paziente schizofrenico OR schizofrenia)

I Gruppo di auto-mutuo aiuto

[C] [Psicoterapia convenzionale OR nessun intervento...]

O Qualità della vita OR stress OR depressione...



Gli operatori booleani devono sempre essere indicati in **MAIUSCOLO** per poter essere interpretati come tali.

In caso contrario vengono letti dai sistemi informatici come “stop words” e di conseguenza ignorati.

Altri elementi di sintassi

Parentesi: raggruppamento algebrico dei concetti
(Migraine OR Headache) AND Zolmitriptan

*** / \$ / ?:** troncamento di un termine
Diabet* ricercherà diabetes, diabetic, diabetology...

Virgolette: ricerca di una frase esatta
“psychiatric rehabilitation”



PRESTAZIONE DELLA QUERY DI RICERCA

Misure di prestazione della query di ricerca

Precisione o **Valore positivo di previsione**
=
Precision or **Positive predictive value**

Richiamo/Recupero o **Sensitività**
=
Recall or **Sensitivity**

Precisione: Definizione

Corrisponde alla proporzione di documenti rilevanti fra tutti quelli recuperati attraverso una stringa di ricerca

$$\frac{\text{n. di doc. rilevanti recuperati}}{\text{n. tot. di doc. recuperati}}$$

ovvero

$$\frac{\text{risultati corretti (veri positivi)}}{\text{risultati corretti + risultati inattesi (veri positivi) + risultati inattesi (falsi positivi)}}$$

Wikipedia, *Information retrieval*, http://it.wikipedia.org/wiki/Information_retrieval [ITA]

Wikipedia, *Precision and recall*, http://en.wikipedia.org/wiki/Precision_and_recall [ENG]

Richiamo/Sensibilità: Definizione

Corrisponde alla proporzione fra il n. di documenti rilevanti recuperati e il n. di tutti i documenti rilevanti disponibili nella collezione considerata (es.: un database)

$$\frac{\text{n. di doc. rilevanti recuperati}}{\text{n. tot. di doc. rilevanti presenti nel DB}}$$

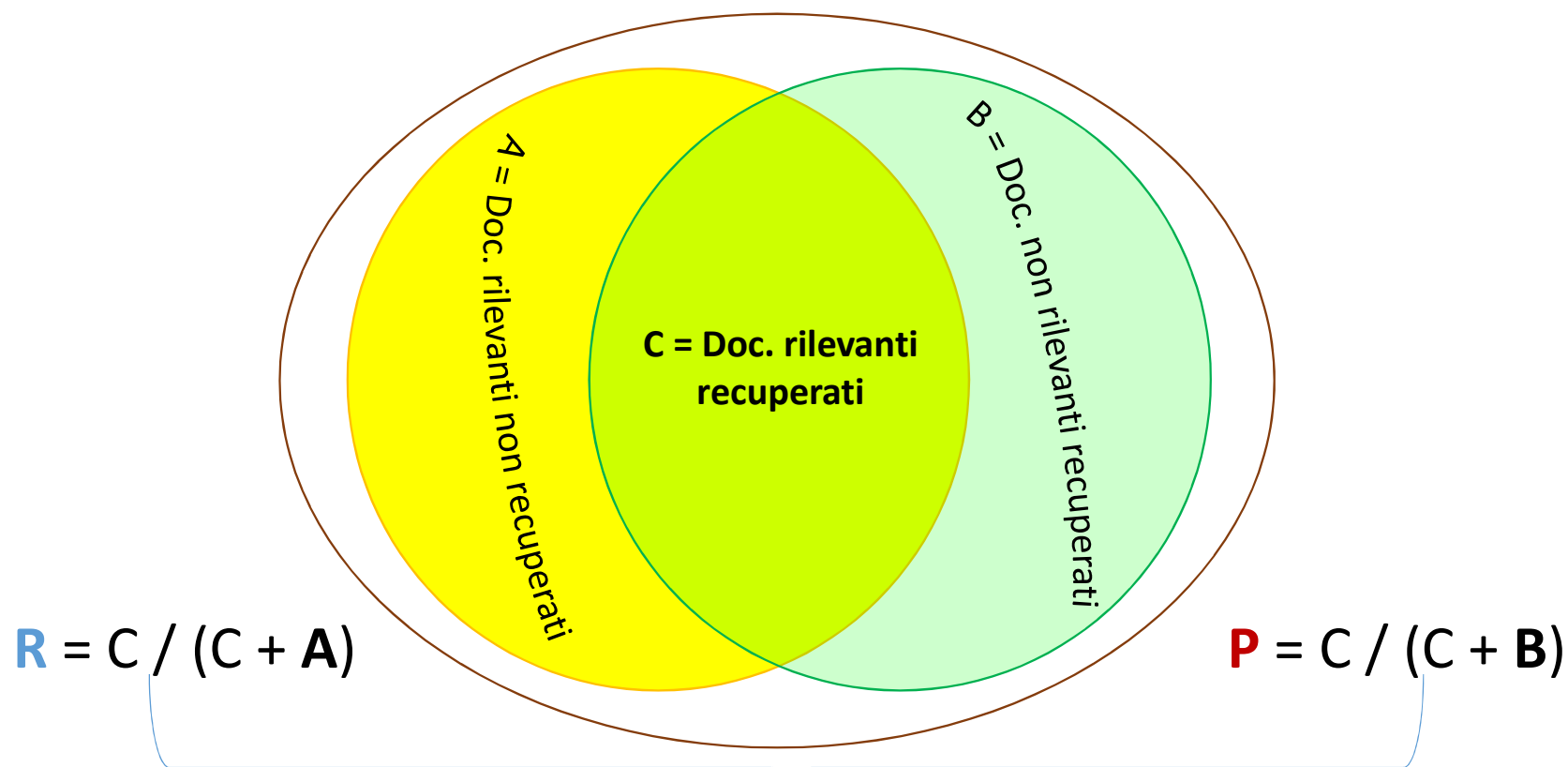
ovvero

$$\frac{\text{risultati corretti (veri positivi)}}{\text{risultati corretti + risultati corretti mancanti (veri positivi) + risultati corretti mancanti (falsi positivi)}}$$

Wikipedia, *Information retrieval*, http://it.wikipedia.org/wiki/Information_retrieval [ITA]

Wikipedia, *Precision and recall*, http://en.wikipedia.org/wiki/Precision_and_recall [ENG]

Richiamo vs Precisione



spesso all'incremento dell'uno corrisponde una diminuzione dell'altro

Precisione vs Richiamo



Eccesso di precisione
Silenzio / Silences

Eccesso di richiamo
Rumore / Noise

“Searches for systematic reviews aim to be as extensive as possible in order to ensure that as many as possible of the necessary and relevant studies are included in the review. It is, however, **necessary to strike a balance between striving for comprehensiveness and maintaining relevance when developing a search strategy.** [...]”

Julian PT Higgins and Sally Green (eds.), *Cochrane Handbook for Systematic Reviews of Interventions*, version 5.1.0, Chap. 6.4.4, <http://www.cochrane-handbook.org>

Precisione, richiamo e booleani

AND e **NOT** agiscono sulla precisione, **OR** sul richiamo.

Per privilegiare il richiamo, un po' a discapito della precisione:

AND

- Evitare un uso eccessivo di concetti diversi in AND.
Limitarsi ai concetti veramente fondamentali!

NOT

- Evitarne l'utilizzo a priori, se non veramente indispensabile.
L'impiego di NOT elimina, infatti, dai risultati anche eventuali studi di confronto che potrebbero invece risultare rilevanti ai fini della ricerca.
Usarlo, piuttosto, a posteriori.

OR

- Utilizzare, per ciascun concetto/elemento del quesito background o del P.I.C.O., una buona varietà di termini associati in OR:
- Termini di tesauruspecifici di una BD bibliografica (es.: MESH) OR parole libere;
 - I **termini di tesauruspecifici** permettono di recuperare automaticamente articoli che usano parole diverse per esprimere lo stesso concetto; forniscono, inoltre, indicazioni di contenuto al di là di quanto semplicemente espresso dal titolo e/o dall'abstract.
 - Utilizzare funzioni specifiche della BD per individuarli. Per ampliarli, controllare anche i termini di tesauruspecifici assegnati a un articolo chiave recuperato e ai suoi **“Related Articles”**.

OR

- Per quanto riguarda le **parole libere**:
 - Associare in OR possibili **sinonimi** (es.: “relatives” OR “caregivers”...);
 - Associare in OR possibili **termini correlati** (es.: “brain” OR “head”...);
 - Associare in OR possibili **varianti di spelling** (es.: “tumor” OR “tumour” / “burnout” OR “burn out”...);
 - Associare in OR possibili **acronimi** (es: “ALS” OR “amyotrophic lateral sclerosis”...).
 - Per ampliarle, controllare le parole utilizzate dagli autori di uno o più articoli chiave recuperati, nonché dai loro **“Related Articles”**.
 - Servirsi del **troncamento (*)**.

TRONCAMENTO (OR)

- L'utilizzo del troncamento genera in automatico una stringa in OR di parole libere accumulate da una radice comune. Es.:

drink* = drink OR drinks OR drinking OR drinker OR drinkers

psychiatr* = psychiatry OR psychiatrist OR psychiatric

- Attenzione! Verificare sempre qual è il simbolo di troncamento utilizzato dalla BD: può, infatti, variare da una piattaforma all'altra (in PubMed, CINAHL e PsycINFO è "*").

bfm@unito.it

www.bfm.unito.it

<https://www.facebook.com/bfm.unito/>